A Generic Framework for Diamond Lemmas

Lars Hellström*

Abstract

This paper gives a generic form of the diamond lemma, which includes support for additive and topological structures of the base set, and which does not require any further structure (e.g. an associative multiplication operation) to be present. This result is intended to be used as the core of diamond lemmas for particular algebraic structures, taking care of all the common technicalities. With this generic diamond lemma, the main steps needed to prove a specialised diamond lemma is to define the reduction maps and analyse the structure of critical ambiguities.

The abstract machinery is backed up with concrete suggestions for how one should set things up in order to reproduce traditional results in the general setting. Several instances of the fundamental theorem of Gröbner basis theory are derived as corollaries of the main result.

1 Introduction

The *Diamond Lemma for Ring Theory* [2] of Bergman is an important theorem that links together several branches of mathematics. On one hand it is the bridge between associative algebra and mathematical logic that can make the definition of an algebra through generators and relations effective. On another it marks a middle ground between the theory of Gröbner bases and the theory of term rewriting, which can be seen as belonging to either of the two. Yet it is only one member in a family of results on similar connections, which can be quite different in their technical details even though the essential ideas are mostly the same. Furthermore many of these results exist in the literature only as sketches (which, it seems, everybody is waiting for someone else to flesh out, as it is all so "obvious" anyway), and as a result the rigor in many arguments becomes somewhat lacking, as they should rightly have been *proofs* making use of some particular diamond lemma.

My intention here is to state and prove a generic form of the diamond lemma from which one can easily derive more specialised results suitable for particular problems. It is probably not the most generic form that is

*E-mail: `Lars.Hellstrom@residenset.net`. Postal address: Lars Hellström, Sand 216, S-881 91 Sollefteå, Sweden.

possible, but it can deal with the technicalities in all cases I know of, and does so without making extensive assumptions about the structure to which it is applied.

From a strictly technical perspective, the theorem given here generalises that of Bergman in three directions:

1. A topological aspect is added to the basic machinery. This makes it possible to treat e.g. formal power series problems within the diamond lemma framework.

2. The assumptions about a multiplicative structure have been dropped from the core theorem. Auxiliary theorems are provided which together with the core theorem cover what the associative algebra diamond lemma can do, but also apply for a much broader range of algebraic structures (nonassociative algebras, operads, PROPs [10], etc.).

3. The definition of reductions has been separated from the diamond lemma, so that it no longer depends on these having a particular form or that all reductions of a particular form are active. The latter is useful if one wishes to cover Shirshov's theory of bases for Lie algebras [14].

The first generalisation was the subject of my Ph.D. thesis [7], but the presentation here has been refined in that it eliminates many minor assumptions on how the multiplicative and topological structures interact. Readers who want concrete examples may however prefer the thesis presentation, as it treats some applications in great detail.

The main advantage of the topological aspect is that it enables one to handle both polynomials and power series (or their respective counterparts from less traditional algebraic structures) using the same machinery. A less apparent advantage is that problems that can be posed entirely in terms of finite sums (i.e., polynomials) sometimes have solutions where the normal form is an infinite sum (power series), and in this case one has to employ the topologized version in order to prove things about this normal form. In so doing, one can take advantage of certain relaxations of the conditions of the classical result; Definition 5.1 of the descending chain condition and Definition 5.9 of ambiguity resolution both admit more than in Bergman's diamond lemma.

The second generalisation has been the main direction in my subsequent work, initiated in response to a question from Loday on whether there is a diamond lemma for operads. It's not too hard to see that there is such a creature — more work had to be spent sorting out the details of claims than the details of their proofs — but one fundamental change when going from algebras to operads is that one goes from a single-sorted algebraic structure (there is *one* set of elements) to a multiple-sorted algebraic structure (in

an operad, elements of different arities don't mix, and hence there is a separate sort of element for each arity). The interactions between these elements of different sorts is certainly a kind of multiplicative structure, but one that syntactically is much more unwieldy (regardless of whether one prefers to phrase it using the structure map formalism or the ith composition formalism) than that of a ring, and bundling these interactions with the diamond lemma would turn an already very technical result into something even worse. Furthermore the generalisations do not stop at operads. There are good reasons to at least go on to PROPs (because each operad is a part of some PROP, and PROPs have a more concise set of axioms), and after that there are more general diagrammatic structures that one may wish to consider. Handling them all in one result does not seem a likely achievement.

What turns out to work is instead to separate the parts of the classical diamond lemma that deal with the multiplicative structure from the parts that ignore this structure. The core of the diamond lemma (Theorem 5.11, with the familiar equivalence of four different conditions) can be very neatly captured as a result on one sort (hence ignoring the multiplicative structure under which sorts may interact), whereas the construction of reductions and verifications that some ambiguities are trivially resolvable fall into the other part. This is actually rather fortunate, because the first part will then deal with the classical technicalities, whereas the second will deal with the particular features of rings, operads, PROPs, or whatever; there is an almost complete separation of responsibilities.

The third generalisation is thus in part a natural consequence of the second, but there are also other advantages to it. One is that many defining identities of the classical nonassociative algebras do not fit well to make "unconditional" rules from; a simple example is the anticommutativity identity of a Lie algebra, which if expressed as a rule '$[x,y] \to -[y,x]$ for all x and y' would lead to the infinite rewrite cycle $[x,y] \to -[y,x] \to -\big(-[x,y]\big) = [x,y] \to -[y,x] \to \cdots$. One way to handle that in practice is to instead make a conditional rule '$[x,y] \to -[y,x]$ if $x > y$' out of it, and the machinery constructed here can handle that; since each pair (x,y) of factors gives rise to a separate reduction map, it is merely a matter of considering only those pairs for which $x > y$ in some suitable ordering of the factors. While there is a certain price to pay in that ambiguity resolution becomes less automatic, this price effectively only comes into play when the conditional rules are involved.

The structure of this paper is as follows. Sections 2 and 3 introduce the framework within which the core diamond lemma (Theorem 5.11) is formulated. These sections also contain plenty of minor constructions for setting up various aspects of this framework, to illustrate features of the formalism used, and to aid the reader in applying the results.

Sections 4 and 5 contain the bulk of the proof of the diamond lemma. The former section is about more abstract algebraic–topological properties of

monoids of maps, whereas the latter introduces an order and uses induction to link these properties to conditions that can be verified through explicit calculations. Notable lemmas are 4.10 (linking normal form uniqueness to univocality of the pointwise limit of reductions), 5.5 (existence of normal forms), and 5.8 (uniqueness of normal form given relative resolvability of ambiguities). Besides the main Theorem 5.11, there is also Theorem 5.6 which provides a characterisation of irreducible elements.

Section 6 is about ambiguities (a.k.a. critical pairs or overlaps) and how one in a multisorted situation can discard non-critical ambiguities from consideration. This is as much about defining 'critical ambiguity' — a subject which spans definitions 6.1, 6.6, and 6.8 — as it is about proving them discardable. The claim that checking the critical ambiguities is as good as checking all ambiguities can be found in Theorem 6.9. Example 6.10 derives Bergman's diamond lemma from the generic theory. Theorem 6.12 is aimed more at completion calculations; it justifies dropping unnecessary rules while in the middle of completing a rewriting system.

Section 7 collects a construction and some technical lemmas that may be used in applications to demonstrate that the most common setting (a collection of free modules) leads to a framework suitable for the generic diamond lemma. Again the aim is to bridge the gap between concrete conditions that are easy to verify and more abstract conditions used in the generic theory.

The final Section 8 is about Gröbner bases, where Theorem 8.4 extends the big equivalence in the generic diamond lemma with some GB-style claims. Several instances of "the fundamental theorem on Gröbner bases" (in commutative, associative, and nonassociative polynomial algebras) are derived as corollaries of this theorem, and the theory is shown to also cover the case of path algebras.

A more practical application of the generic diamond lemma theory can be found in [8]. Unlike the applications in Section 8, this exercises the multisorted aspects of the framework.

Notation

The set $\mathbb{N}$ of natural numbers is considered to include 0. $\mathbb{Z}^+$ is the set of positive integers and $\mathbb{R}^+$ is the set (sometimes the multiplicative group) of positive real numbers. The shorthand $f(A)$ for $\{\, f(a) \mid a \in A \,\}$ is frequently applied.

Formal variables are typically written using a sans-serif font: a, b, c, etc. When X is a set of such letters, X^* denotes the free monoid on X, i.e., the set of all finite strings of elements from X. The identity element in X^* is denoted $\mathsf{1}$.

On the matter of monomials versus terms, a *monomial* is considered to not include a coefficient, whereas a *term* generally contains a coefficient. The relation symbol $\equiv$ denotes congruence rather than identity.

2 Basics

For the machinery employed here, it is convenient to fix a framework with five pieces of data:

- An abelian group $\mathcal{M}$ (written additively). This will play the role of set of all finite expressions.
- A set R of maps $\mathcal{M} \longrightarrow \mathcal{M}$. This can be used to encode a module structure on $\mathcal{M}$.
- A subset $\mathcal{Y}$ of $\mathcal{M}$. This will play the role of set of monomials.
- A family $\mathcal{O} = \{B_n\}_{n=1}^{\infty}$ of subsets of $\mathcal{M}$. This will become the fundamental system of neighbourhoods of $0 \in \mathcal{M}$ and is thus defining the topology.
- A family $T_1(S)$ of maps $\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$, where $\overline{\mathcal{M}} \supseteq \mathcal{M}$ is the set of all expressions. These are what in the end specify the wanted congruence on $\mathcal{M}$.

When applying the diamond lemma to a multiple-sorted structure, there will be one such quintuplet $\big(\mathcal{M}, R, \mathcal{Y}, \mathcal{O}, T_1(S)\big)$ for each sort, but since the core diamond lemma itself is applied separately for each sort, one does not have to take this multiplicity into account when proving it. Notation for and interactions between different framework quintuplets for a structure are considered in Section 6.

In the main theorem there will also be:

- a partial order P on $\mathcal{Y}$;

but that can without too much difficulty be separated from the rest of the machinery, so it will instead be introduced explicitly whenever it is needed. Having it separate is sometimes convenient, as one in complicated arguments might want to make use of several different orders. Finally, there is in several supporting results:

- a family V of maps $\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$, which can be used to enforce compatibility with a multiplicative structure;

but the typical use of that item is rather on the level of constructing $T_1(S)$ or proving things about it.

The choices of R, $\mathcal{Y}$, $\mathcal{O}$, and $T_1(S)$ are subject to a couple of additional conditions, which are specified as *assumptions* below. Technically it would be possible to instead include them as additional conditions in all theorems and lemmas that depend on them, but it is more convenient to throughout the presentation assume them to be satisfied. There are plenty of suggestions for how one may choose the framework data to ensure that the assumptions

are met. $T_1(S)$ is treated in the next section, but assumptions on R, $\mathcal{Y}$, and $\mathcal{O}$ are given here.

Assumption 1. *Every element of R is a group endomorphism of $\mathcal{M}$.*

Definition 2.1. A subgroup $N \subseteq \mathcal{M}$ is said to be an **R-module** if $r(a) \in N$ for all $a \in N$ and $r \in R$.

If $\mathcal{M}$ has an $\mathcal{R}$-module structure for some ring $\mathcal{R}$, then it is natural to choose as R the set of maps $a \mapsto ra : \mathcal{M} \longrightarrow \mathcal{M}$ for all $r \in \mathcal{R}$; in this case the above R-module concept coincides with the standard $\mathcal{R}$-module concept and all is as one expects it to be. It may however in some cases be necessary to impose a restriction on the elements of $\mathcal{R}$ which may contribute to R, and in that case it is weaker to be an R-module than to be an $\mathcal{R}$-module. It is also perfectly possible to take $R = \varnothing$ if no particular module structure is available.

Assumption 2. *If $N \subseteq \mathcal{M}$ is an R-module such that $\mathcal{Y} \subseteq N$ then $N = \mathcal{M}$.*

In other words, $\mathcal{Y}$ spans $\mathcal{M}$. The traditional approach is to begin with $\mathcal{Y}$ just being some set, pick some ring $\mathcal{R}$, and then *construct* $\mathcal{M}$ as the free $\mathcal{R}$-module with basis $\mathcal{Y}$; if in particular $\mathcal{Y}$ is the set X^* of words on the alphabet X then this will make $\mathcal{M}$ equal to the free $\mathcal{R}$-algebra $\mathcal{R}\langle X\rangle$. An alternative approach for the free algebra $\mathcal{R}\langle X\rangle$ is however to let $\mathcal{Y}$ be the set of all terms — products $r\mu$ of a scalar $r \in \mathcal{R}$ and a monomial $\mu \in X^*$ — as this makes it possible to take $R = \varnothing$. It is also possible to interpolate between these two extremes, or pick a set $\mathcal{Y}$ with more complicated linear dependencies between elements, although the latter is likely to make it more complicated to construct $T_1(S)$.

Definition 2.2. Let R^* denote the set of all finite compositions of elements of R; in particular, R^* is considered to contain the identity map $\mathrm{id}\colon \mathcal{M} \longrightarrow \mathcal{M}$. Let $\pm R^*$ denote the set $\{\, r, -r \,|\, r \in R^* \,\}$. Let $R^*\mathcal{Y}$ denote the set $\{\, r(\mu) \,|\, r \in R^*, \mu \in \mathcal{Y} \,\} \subseteq \mathcal{M}$.

Lemma 2.3. *Every element $a \in \mathcal{M}$ can be expressed as*

$$a = \sum_{k=1}^{n} r_k(\mu_k) \tag{2.1}$$

for some $n \in \mathbb{N}$, $r_1, \ldots, r_n \in \pm R^*$, and $\mu_1, \ldots, \mu_n \in \mathcal{Y}$.

Proof. The set of elements on the form (2.1) constitutes an R-module that contains $\mathcal{Y}$. Hence by Assumption 2 the set of such elements is the whole of $\mathcal{M}$. □

Assumption 3. *$\mathcal{O} = \{B_n\}_{n=1}^{\infty}$ is a family of R-modules such that $B_n \supseteq B_{n+1}$ for all $n \in \mathbb{Z}^+$ and $\bigcap_{n=1}^{\infty} B_n = \{0\}$.*

Definition 2.4. A set $N \subseteq \mathcal{M}$ is said to be **open** (in $\mathcal{M}$) if there for every $a \in N$ exists some $\varepsilon \in \mathcal{O}$ such that

$$N \supseteq \{\, a + b \,\big|\, b \in \varepsilon \,\}. \tag{2.2}$$

To put it differently: The topology on $\mathcal{M}$ is the group topology for which $\mathcal{O}$ is a fundamental system of neighbourhoods of 0.

Many arguments involving topology in subsequent sections will be expressed using ε-δ-formalism, but since ε and δ will be neighbourhoods of 0 rather than the conventional positive real numbers, a few examples of what this formalism looks like may be in order. First and foremost, the *ε-neighbourhood* of an element a is the set $a + \varepsilon := \{\, a + b \,|\, b \in \varepsilon \,\}$. A map f is continuous at 0 if there for every $\varepsilon \in \mathcal{O}$ exists some $\delta \in \mathcal{O}$ such that $f(\delta) := \left\{\, f(a) \,\middle|\, a \in \delta \,\right\} \subseteq \varepsilon$. It should furthermore be observed that continuity at 0, for group homomorphisms, is equivalent to continuity everywhere (and even to uniform continuity everywhere). That δ is smaller than (or equal to) ε is of course expressed as $\delta \subseteq \varepsilon$, and the minimum of ε and δ is $\varepsilon \cap \delta$.

In addition to these general properties of neighbourhood arithmetic, there are also some special properties following from Assumption 3 that are of great importance here. Firstly $\varepsilon + \varepsilon = \varepsilon$ for every $\varepsilon \in \mathcal{O}$ since ε is a group. Similarly $\varepsilon - \varepsilon = \varepsilon$ and $r(\varepsilon) \subseteq \varepsilon$ for all $r \in R$. Finally the inclusion of B_m in B_n whenever $m > n$ implies that $\varepsilon + \delta = \varepsilon \cup \delta$ for all $\varepsilon, \delta \in \mathcal{O}$.

The choice of $\mathcal{O}$ is a rather extensive topic, with many different approaches that should be mentioned, so it seems best to leave that for the end of this section, and instead proceed with the things that can be done as soon as $\mathcal{O}$ is in place.

Lemma 2.5. *The group operations on $\mathcal{M}$ and all elements in R are continuous.*

Proof. Let $s\colon \mathcal{M} \times \mathcal{M} \longrightarrow \mathcal{M} : (a, b) \mapsto a - b$ be subtraction as a map; proving it continuous implies the same for the standard group operations addition and negation. Let $N \subseteq \mathcal{M}$ be an arbitrary open set and let $(a, b) \in s^{-1}(N)$ be arbitrary too. Since $a - b \in N$ there exists some $\varepsilon \in \mathcal{O}$ such that $(a-b)+\varepsilon \subseteq N$. For this ε, $(a+\varepsilon)-(b+\varepsilon) = (a-b)+(\varepsilon-\varepsilon) = (a-b)+\varepsilon \subseteq N$, and hence $(a+\varepsilon) \times (b+\varepsilon) \subseteq s^{-1}(N)$, which means (a, b) is an interior point of $s^{-1}(N)$. It follows that s^{-1} maps open sets to open sets, and hence s is continuous.

Now let $r \in R$ and an open set $N \subseteq \mathcal{M}$ be arbitrary. For every $a \in r^{-1}(N)$ there exists some $\varepsilon \in \mathcal{O}$ such that $r(a) + \varepsilon \subseteq N$, and hence $a + \varepsilon \subseteq r^{-1}(N)$ because $r(a + \varepsilon) = r(a) + r(\varepsilon) \subseteq r(a) + \varepsilon \subseteq N$. Thus r is continuous. □

The next step is to go from $\mathcal{M}$ to its completion $\overline{\mathcal{M}}$, which can be constructed in the standard way as the set of equivalence classes of Cauchy sequences in $\mathcal{M}$, where two sequences $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$ are considered equivalent if $\lim_{n\to\infty}(a_n - b_n) = 0$. The topology in the completion can be defined in terms of limits: $a = \lim_{n\to\infty} a_n$ for $a_n = \left[\{b_{n,k}\}_{k=1}^\infty\right]$ if and only if $\{b_{n,n}\}_{n=1}^\infty$ is a Cauchy sequence in $\mathcal{M}$ and $a = \left[\{b_{n,n}\}_{n=1}^\infty\right]$. The equivalence classes of the constant sequences provide the canonical embedding of $\mathcal{M}$ into its completion, and it is convenient to identify this with the original $\mathcal{M}$.

An alternative approach, which fits better in with many textbook definitions of the completion, is to turn $\mathcal{M}$ into a metric space and make use of this explicit metric when defining e.g. the topology of $\overline{\mathcal{M}}$. (Both approaches yield the same end result.) There are several metrics which all reproduce the topology of $\mathcal{M}$, but the following is often the most natural:

$$d(a,b) = \begin{cases} 1 & \text{if } a - b \notin B_1, \\ \inf\{\, 2^{-n} \,\big|\, a - b \in B_n \,\} & \text{otherwise} \end{cases} \tag{2.3}$$

for all $a, b \in \mathcal{M}$. That $d(a,a) = 0$ makes use of the infimum, whereas in the formula for $d(a,b)$ when $a \neq b$ this inf is equivalent to a min.

Definition 2.6. The completion of $\mathcal{M}$ is denoted $\overline{\mathcal{M}}$, and $\mathcal{M}$ is considered to be a subset of $\overline{\mathcal{M}}$. For any $N \subseteq \overline{\mathcal{M}}$, the topological closure in $\overline{\mathcal{M}}$ of N is denoted $\overline{N}$. Let $\widehat{\mathcal{O}} = \left\{\overline{B_n}\right\}_{n=1}^\infty$.

The group operations extend by continuity to the whole of $\overline{\mathcal{M}}$, as do the homomorphisms in R, and will henceforth be considered to be defined on the whole of $\overline{\mathcal{M}}$. Accordingly, any subgroup $N \subseteq \overline{\mathcal{M}}$ is said to be an **R-module** for which $r(a) \in N$ whenever $a \in N$ and $r \in R$. If $Z \subseteq \overline{\mathcal{M}}$ is some set, then $\mathrm{Span}(Z)$ will denote the smallest R-module which contains Z. Denote by $\mathrm{Cspan}(Z)$ the topological closure of $\mathrm{Span}(Z)$.

Lemma 2.7. *$\widehat{\mathcal{O}}$ is a fundamental system of neighbourhoods of 0 in $\overline{\mathcal{M}}$. In particular the elements of $\widehat{\mathcal{O}}$ are clopen (simultaneously closed and open), whence the topology of $\overline{\mathcal{M}}$ is zero-dimensional and totally disconnected.*

Proof. First consider an arbitrary $F \subseteq \overline{\mathcal{M}}$ that is closed and not disjoint from $\overline{B_n}$ for any $n \in \mathbb{Z}^+$; as an auxiliary result it will be shown that such an $F \ni 0$. The closure $\overline{N}$ of some $N \subseteq \mathcal{M}$ consists of those points $a \in \overline{\mathcal{M}}$ for which there exists some Cauchy sequence $\{a_k\}_{k=1}^\infty \subseteq N$ such that $a = \lim_{k\to\infty} a_k$. Let $\{a_n\}_{n=1}^\infty \subseteq \overline{\mathcal{M}}$ be a sequence such that $a_n \in F \cap \overline{B_n}$ and let $\{b_{n,m}\}_{m,n=1}^\infty \subseteq \mathcal{M}$ be a collection of points such that $a_n = \lim_{m\to\infty} b_{n,m}$ and $\{b_{n,m}\}_{m=1}^\infty \subseteq B_n$ for all n. Clearly $b_{n,n} \in B_n$ for all n and thus $b_{n,n} \to 0$ as $n \to \infty$, which implies $\lim_{n\to\infty} a_n = 0$ as well. Since F was closed, it follows that $0 \in F$.

Now let $U \subseteq \overline{\mathcal{M}}$ be an arbitrary open neighbourhood of 0, and consider the matter of whether U contains some $\overline{B_n}$ for n large enough. The complement $F = \overline{\mathcal{M}} \setminus U$ is a closed set that does not contain 0, and hence by the

converse of the above result there is some n for which $F \cap \overline{B_n} = \varnothing$, meaning that $\overline{B_n} \subseteq U$.

Consider next the problem of showing that all the $\overline{B_n}$ are clopen (both open and closed). A Cauchy sequence $\{a_k\}_{k=1}^{\infty} \subseteq \mathcal{M}$ has the property that it is *either* eventually in B_n *or* eventually in the complement $\mathcal{M} \setminus B_n$, because by the definition of Cauchy sequence there exists some m such that if $i, j \geqslant m$ then $a_i - a_j \in B_n$, or equivalently $a_i \in a_j + B_n$, and thus if $a_j \in B_n$ for some $j \geqslant m$ then $a_i \in a_j + B_n \subseteq B_n + B_n = B_n$ for all $i \geqslant m$, in which case $\{a_k\}_{k=1}^{\infty}$ indeed is eventually in B_n. If no $a_j \in B_n$ for $j \geqslant m$ then instead $a_j \in \mathcal{M} \setminus B_n$ for all $j \geqslant m$, and consequently $\{a_k\}_{k=1}^{\infty}$ will be eventually in $\mathcal{M} \setminus B_n$. This property of Cauchy sequences means a sequence can converge either to an element of $\overline{B_n}$ or to an element of $\overline{\mathcal{M} \setminus B_n}$, but not both, and therefore these sets will be disjoint; $\overline{B_n} \cup \overline{\mathcal{M} \setminus B_n}$ is a partition of $\overline{\mathcal{M}}$. Since both parts in this partition are closed by definition, they are also both open, and in particular $\overline{B_n}$ is both closed and open.

This has shown that $\widehat{\mathcal{O}}$ is a fundamental system of neighbourhoods of 0 in $\overline{\mathcal{M}}$, and that its members are all clopen. A topology is said to be zero-dimensional if it has a basis consisting entirely of clopen sets, and every space with a zero-dimensional topology is totally disconnected. □

Zero-dimensional topologies are, just like Zariski topologies, perfectly fine topologies (i.e., all the axioms hold and hence the basic theorems follow), but a bit unsettling when one first encounters them as things do not behave in quite the way one has gotten used to in the standard topology on $\mathbb{R}$ — the multitude of sets that are open *and* closed at the same time being the most obvious oddity. Being metric, and consequently Hausdorff, the topology on $\overline{\mathcal{M}}$ does however have much more in common with the standard topology on $\mathbb{R}$ than it has with Zariski topologies, so it is not all that far out.

Exercise. An intuition for spaces like $\overline{\mathcal{M}}$ may be found by comparing them to Cantor sets, as the two have many traits in common. Indeed, for $\mathcal{M} = \mathbb{Z}_2[x]$ (univariate polynomials over $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$) and $B_n = \mathcal{M}x^{\lceil \alpha n \rceil}$ where $\alpha = \log_3 2$, the completion $\overline{\mathcal{M}}$ is very similar to the standard Cantor set. Show that the map $\phi\colon \mathcal{M} \longrightarrow [0,1]$ defined by

$$\phi\biggl(\sum_{k=0}^{n}(s_k + 2\mathbb{Z})x^k\biggr) = \frac{2}{3}\sum_{k=0}^{n} s_k 3^{-k} \qquad \text{for all } \{s_k\}_{k=0}^{n} \subseteq \{0,1\}$$

satisfies

$$C_1 d(a,b) \leqslant \bigl|\phi(a) - \phi(b)\bigr| \leqslant C_2 d(a,b) \quad \text{for all } a, b \in \mathcal{M},$$

for some positive constants C_1 and C_2, where $|\cdot|$ is the standard absolute value on $\mathbb{R}$ and d is the metric from (2.3). Conclude that ϕ extends to a homeomorphism from $\overline{\mathcal{M}}$ to the remove-middle-third Cantor set on the unit interval.

Lemma 2.8. *If $A \subseteq \overline{\mathcal{M}}$ and $a \in \mathrm{Cspan}(A)$ then for every $\varepsilon \in \widehat{\mathcal{O}}$ there exists a natural number n, some elements $\{a_i\}_{i=1}^n \subseteq A$, and some maps $\{r_i\}_{i=1}^n \subseteq \pm R^*$ such that*

$$\sum_{i=1}^{n} r_i(a_i) \in a + \varepsilon. \tag{2.4}$$

Proof. By definition of topological closure applied to $\mathrm{Cspan}(A)$, there exists some $b \in \mathrm{Span}(A)$ such that $a - b \in \varepsilon$. Since the set of all elements on the form $\sum_{i=1}^n r_i(a_i)$ for $\{a_i\}_{i=1}^n \subseteq A$ and $\{r_i\}_{i=1}^n \subseteq \pm R^*$ constitute an R-module containing A, it follows that b has an expression on that form. □

For $A = \mathcal{Y}$, this lemma is a topologized version of Lemma 2.3, but Lemma 3.7, Definition 5.3, Theorem 5.6, and Definition 5.7 all characterise important subsets of $\overline{\mathcal{M}}$ as being on the form $\mathrm{Cspan}(A)$ for a suitable $A \subseteq \overline{\mathcal{M}}$.

The rest of this section is a discussion of some important methods for constructing a topology on $\mathcal{M}$, i.e., for choosing a system of neighbourhoods $\mathcal{O}$. The trivial choice is to let $B_n = \{0\}$ for all n; this equips $\mathcal{M}$ with the discrete topology, makes $\overline{\mathcal{M}} = \mathcal{M}$, and simplifies the machinery below quite considerably. This is also the choice one should use if one wishes to reproduce Bergman's diamond lemma.

A nontrivial choice of topology which has long traditions in algebra is that of an ideal-adic topology. In this case it is assumed that $\mathcal{M}$ also has a multiplicative structure, and as B_1 is chosen a nontrivial ideal in $\mathcal{M}$. Then each B_n is defined as the nth ideal power B_1^n of B_1, i.e., the ideal generated by all products of n elements from B_1. Such choices of $\mathcal{O}$ allow localisations of $\mathcal{M}$ to be treated within this framework. The condition that all these B_n are R-modules is not necessarily fulfilled for this construction, but it follows very naturally when for example $\mathcal{M}$ is an $\mathcal{R}$-algebra and R is the set of multiplication-by-a-scalar maps. Nor is necessarily $\bigcap_{n=1}^{\infty} B_1^n = \{0\}$ for every ideal B_1, but it typically holds for the interesting choices of B_1.

A generalisation of the class of ideal-adic topologies is provided by the 'filtered structures' of Robbiano [13] and Mora [11]. Here it is again assumed that $\mathcal{M}$ is a ring, and a Γ-filtration $\{F_\gamma\}_{\gamma\in\Gamma}$ of $\mathcal{M}$ is given. This means Γ is assumed to be a totally ordered semigroup (written additively, but at least in [11] not assumed to be commutative), and the F_γ are assumed to be subgroups of $\mathcal{M}$ which satisfy:

(R1) If $\gamma, \delta \in \Gamma$ are such that $\gamma < \delta$ then $F_\gamma \subseteq F_\delta$.

(R2) $F_\gamma \cdot F_\delta \subseteq F_{\gamma+\delta}$ for all $\gamma, \delta \in \Gamma$.

(R3) For every $a \in \mathcal{M} \setminus \{0\}$ the set $\{\, \gamma \in \Gamma \,\big|\, F_\gamma \ni a \,\}$ has a minimal element.

For actual results, these authors typically also assume that Γ is 'inf-limited', which means that for any infinite strictly descending sequence $\gamma_1 > \gamma_2 > \gamma_3 > \cdots$ in Γ and any given $\gamma \in \Gamma$, there exists an n such that $\gamma_n < \gamma$. In this case, one can simply choose one such infinite strictly descending sequence $\gamma_1 > \gamma_2 > \gamma_3 > \cdots$ in Γ and define $\mathcal{O}$ by letting $B_n = F_{\gamma_n}$ for all $n \in \mathbb{Z}^+$; it follows from (R3) and inf-limitedness that $\bigcap_{n=1}^{\infty} B_n = \{0\}$. Also observe that the resulting topology on $\mathcal{M}$ is the same regardless of which sequence $\{\gamma_n\}_{n=1}^{\infty}$ is chosen.

The abstract setting of a filtered structure only supports setting $R = \varnothing$, but again the F_γ are in many concrete cases modules over a ring of scalars, and then it is possible to encode the whole of that ring into R. Otherwise it is a rather striking feature of the filtered structure machinery that one does not assume any "coefficients" to exist from start, but rather constructs them from the filtered structure. Defining

$$\begin{aligned} V_\gamma &:= \bigcup_{\substack{\delta \in \Gamma \\ \delta < \gamma}} F_\delta, \\ G_\gamma &:= F_\gamma / V_\gamma, \\ G &:= \bigoplus_{\gamma \in \Gamma} G_\gamma \end{aligned}$$

one gets the associated graded ring G that can be used as a coordinatized form of $\mathcal{M}$. Each coordinate a_γ then assumes values in the corresponding G_γ; these groups may vary quite a lot in size and structure, but for reasonable cases of $\mathcal{M}$ being an $\mathcal{R}$-algebra it often holds that each G_γ is as a group isomorphic to either $\mathcal{R}$ or $\{0\}$. It is also common that a filtered structure is equipped with a map $f\colon \Gamma \longrightarrow \mathcal{M}$ such that $f(\gamma) \in F_\gamma$ for all $\gamma \in \Gamma$ (and $f(\gamma) \notin V_\gamma$ whenever $V_\gamma \neq F_\gamma$), although that map is not part of the formal definition. The image of such an f is typically the primary candidate for $\mathcal{Y}$, even though there is nothing in the generic formalism from which one may deduce that this image should span $\mathcal{M}$.

Even more general is the approach to define $\mathcal{O}$ as a family of balls with respect to an ultranorm $\|\cdot\|$ on $\mathcal{M}$:

$$B_n = \left\{ a \in \mathcal{M} \,\middle|\, \|a\| < 2^{-n} \right\} \qquad \text{for all } n \in \mathbb{Z}^+. \tag{2.5}$$

In one sense this construction is universal, because if $\mathcal{O} = \{B_n\}_{n=1}^{\infty}$ is given then one can always use (2.3) to reconstruct a norm $\|a\| = d(a, 0)$ that in (2.5) would give rise to the original neighbourhood system $\mathcal{O}$, but more important is that it often provides a convenient method for arriving at a topology with desirable properties.

The standard construction of an ultranorm in the case $\mathcal{M} = \mathcal{R}\langle X \rangle$ is to pick any function $w\colon X \longrightarrow \mathbb{R}$ and define the ultranorm $\|\cdot\|$ on $\mathcal{Y} = X^*$ to be the unique monoid homomorphism $\mathcal{Y} \longrightarrow \mathbb{R}^+$ that satisfies $\|x\| = 2^{w(x)}$ for

all $x \in X$; in other words $\|\prod_{i=1}^{n} x_i\| := \prod_{i=1}^{n} 2^{w(x_i)}$ for all $x_1 x_2 \cdots x_n \in \mathcal{Y}$. This is then extended to the whole of $\mathcal{M}$ by defining

$$\|a\| := \max_{\mu \in Z} \|\mu\| \quad \text{where } Z \subset \mathcal{Y} \text{ is minimal such that } a \in \mathrm{Span}(Z), \quad (2.6)$$

and in particular letting $\|0\| := 0$, as a sort of $0 = \max \varnothing$. The w is known as the *weight function* for the norm, and its sign determines how the formal variables behave; if $w(x) \geqslant 0$ then x will be a polynomial-style variable, whereas if $w(x) < 0$ then x will be a power-series-style variable. The logarithm of $\|\cdot\|$ behaves as a weighted polynomial-style degree function on $\mathcal{M} = \mathcal{R}\langle X \rangle$.

Definition 2.9. Formally, a function $a \mapsto \|a\| : \mathcal{M} \longrightarrow \mathbb{R}$ is said to be a (group) **ultranorm** if

(i) $\|a\| \geqslant 0$ for all $a \in \mathcal{M}$.

(ii) $\|a - b\| \leqslant \max\{\|a\|, \|b\|\}$ for all $a, b \in \mathcal{M}$.

(iii) $\|a\| = 0$ for some $a \in \mathcal{M}$ if and only if $a = 0$.

If $\mathcal{M}$ is a ring and in addition

(iv) $\|ab\| \leqslant \|a\| \, \|b\|$ for all $a, b \in \mathcal{M}$

then $\|\cdot\|$ is said to be a **ring ultranorm**. If instead $\mathcal{R}$ is a ring with ultranorm $|\cdot|$ and $\mathcal{M}$ is an $\mathcal{R}$-module, then a group ultranorm $\|\cdot\|$ on $\mathcal{M}$ is said to be a **module ultranorm** if

(v) $\|ra\| \leqslant |r| \, \|a\|$ for all $r \in \mathcal{R}$ and $a \in \mathcal{M}$.

An **algebra ultranorm** has to satisfy all of (i)–(v).

The 'ultra' prefix pertains primarily to property (ii) — the *strong triangle inequality* — and in particular to its right hand side $\max\{\|a\|, \|b\|\}$, which is more strict than the $\|a\| + \|b\|$ of the ordinary triangle inequality. Among the direct consequences of (ii) are that any ε-neighbourhood of 0 — i.e., any set of the form $\{\, a \in \mathcal{M} \mid \|a\| < \varepsilon \,\}$ for a real number $\varepsilon > 0$ — is a subgroup of $\mathcal{M}$.

The *trivial ultranorm* has $\|0\| = 0$ and $\|a\| = 1$ for all $a \neq 0$; it exists for all groups and reproduces the discrete topology.

If $\mathcal{M}$ is an $\mathcal{R}$-module and the ultranorm $\|\cdot\|$ on $\mathcal{M}$ satisfies (2.6) then equipping $\mathcal{R}$ with the trivial ultranorm will make $\|\cdot\|$ an $\mathcal{R}$-module ultranorm. This is typically the "correct" scalar norm for a formal power series problem, as all nonzero scalar values are then equivalent for matters of series convergence. Conversely it is often convenient to define the norm on $\mathcal{M}$ so that it becomes an $\mathcal{R}$-module norm with respect to some given norm $|\cdot|$ on $\mathcal{R}$. In the particular cases where $\mathcal{Y}$ is an $\mathcal{R}$-module basis for $\mathcal{M}$, then one

may choose to make $\|r\mu\| = |r| \, \|\mu\|$ for all $r \in \mathcal{R}$ and $\mu \in \mathcal{Y}$, which has its advantages when it comes to defining $T_1(S)$ below. Non-trivial scalar norms may however require nontrivial choices also of R and $\mathcal{Y}$.

What complicates the choice of a scalar norm is the assumption that each B_n is an R-module, which given (2.5) is equivalent to the claim that $\big\|r(a)\big\| < 2^{-n}$ for all $r \in R$ and $a \in \mathcal{M}$ satisfying $\|a\| < 2^{-n}$. One would typically ensure this by enforcing the condition that $\big\|r(a)\big\| \leqslant \|a\|$ for all $r \in R$ and $a \in \mathcal{M}$, and this will indeed be the case provided $r \colon \mathcal{M} \longrightarrow \mathcal{M}$ is a map on the form $a \mapsto sa$ for some scalar $s \in \mathcal{R}$ such that $|s| \leqslant 1$, as then $\big\|r(a)\big\| = \|sa\| \leqslant |s| \, \|a\| \leqslant \|a\|$. Hence the natural choice of R when $\mathcal{M}$ is an $\mathcal{R}$-module with a ditto ultranorm is to take

$$R = \big\{ \, a \mapsto ra : \mathcal{M} \longrightarrow \mathcal{M} \, \big| \, r \in \mathcal{R} \text{ and } |r| \leqslant 1 \, \big\} \, . \tag{2.7}$$

Unless all scalars $r \in \mathcal{R}$ satisfy $|r| \leqslant 1$, this will make the R-module concept distinct from that of an $\mathcal{R}$-module however, and this has repercussions elsewhere. $\mathcal{Y}$ must span $\mathcal{M}$ as an R-module, so if Y is an $\mathcal{R}$-module basis of $\mathcal{M}$ then $\mathcal{Y}$ may have to be chosen as something like the set of all products $r\mu$ for $r \in \mathcal{R}$ and $\mu \in Y$ in order to make it all fit.

A case where this predicament arises is that of $\mathcal{M}$ being a module over the p-adic numbers $\mathbb{Q}_p$, as these come equipped with an ultranorm (the p-adic valuation) that has $|p^n| = p^{-n}$ for all $n \in \mathbb{Z}$. The R defined by (2.7) for $\mathcal{R} = \mathbb{Q}_p$ is isomorphic to the p-adic integers, but not to the entire field of p-adic numbers; conversely any B_n defined by (2.5) will fail to be closed under multiplication by the scalar p^{-1} and is thus not a $\mathbb{Q}_p$-module although it will be an R-module. A $\mathbb{Q}_p$-module basis Y for $\mathcal{M}$ will in this case not be large enough to serve as $\mathcal{Y}$. One can instead use the set of all terms as suggested above, but since p^{-1} together with the p-adic integers generate the whole of $\mathbb{Q}_p$, it is also sufficient to make $\mathcal{Y}$ the set of all products $p^{-n}\mu$ for $\mu \in Y$ and $n \in \mathbb{Z}$ (or even $n \in \mathbb{N}$). It is typically easier to construct the partial order P on $\mathcal{Y}$ if the latter has a simple, discrete structure.

3 Reductions

While the purpose of introducing $\mathcal{M}$, R, $\mathcal{Y}$, and $\mathcal{O}$ is primarily to fix and structure a stage for the diamond lemma, $T_1(S)$ is what provides the actors on that stage, so that a play of equivalence and normal forms may be performed. The relation of $T_1(S)$ to the equivalence of elements in $\overline{\mathcal{M}}$ is primarily that these maps preserve it — $t(a)$ must be equivalent to a for all $t \in T_1(S)$ and $a \in \overline{\mathcal{M}}$ — but as the equivalence concept is derived from $T_1(S)$, this is a theorem rather than an assumption. The role of the collection $T_1(S)$ of maps $\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ is mostly that of a presentation: it is not uniquely determined by that which it is used to define, but it gets the job done, and you may use it to verify at least some conditions about the whole.

In applications one rarely starts from $T_1(S)$ — hence the somewhat odd notation; $T_1(S)$ is typically constructed from a more fundamental set S of directed equivalences, a so-called *rewriting system* — but for this general proof it provides the best balance between abstract adaptability and concrete constructibility. As in the previous section some examples will be given below of how $T_1(S)$ can be constructed from such a more fundamental S, but this should be taken more as hints than as a full survey; it is sometimes necessary to combine several different methods of construction. That applications typically make some S the fundamental entity has however influenced the choice of notations below, in that every object that *formally* depends on the choice of $T_1(S)$ is *written* as though it would depend on 'S'. Besides being more convenient in applied instances of the diamond lemma, this choice of notations also simplifies comparisons with [7].

Definition 3.1. Let $T(S)$ be the set of all finite compositions of elements from $T_1(S)$; in particular the identity map $\mathrm{id}\colon \overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ is considered to be an element of $T(S)$, on account of being the composition of an empty sequence of maps in $T_1(S)$. The elements of $T(S)$ are called **reductions** and the elements of $T_1(S)$ in particular are called **simple reductions**.

The name *reduction* suggests that these maps take something away, and this is indeed typically the case. Standard constructions of reductions tend to make them more or less projections, and although there is no formal need for them to be, it may be helpful on a first reading to think of them that way. It should however be observed that even those reductions which really are projections tend to be rather skew and have very small kernels, so don't expect to use just one and be done with it; getting anywhere is much more like a round of golf, where one has to hit the ball repeatedly (typically making use of many different clubs) in order to get it into the hole.

The distinction between simple and non-simple reductions is mostly in the eye of the beholder, because nothing prevents picking as $T_1(S)$ a set of maps that constitute a monoid under composition, in which case one would have $T(S) = T_1(S)$ and all reductions would be simple. The point of letting the user designate some reductions as being simple is that it is often sufficient to verify a condition only for the simple ones, as the property in question easily extends to all reductions.

Assumption 4. *Every simple reduction $t \in T_1(S)$ is a continuous group homomorphism $\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ which satisfies $t \circ r = r \circ t$ for all $r \in R$.*

Reductions may of course satisfy this property for a larger class of maps than R; they may for example all be $\mathcal{R}$-linear for some ring $\mathcal{R}$ that is larger than R. Therefore many lemmas below that say some set is an R-module will have a 'more generally, ...' clause which covers the case of additional maps r that commute with all reductions.

Lemma 3.2. *Every reduction $t \in T(S)$ is a continuous group homomorphism $\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ which satisfies $t \circ r = r \circ t$ for all $r \in R^*$.*

Proof. The identity map $\mathrm{id} \in T(S)$ trivially satisfies the properties in the lemma. All other reductions are finite compositions of simple reductions, and since the composition of two continuous group homomorphisms is a continuous group homomorphism, it follows from Assumption 4 that all reductions are continuous group homomorphisms. Finally if $t = t_n \circ \cdots \circ t_1$ for $t_1, \ldots, t_n \in T_1(S)$ and $r = r_m \circ \cdots \circ r_1$ for $r_1, \ldots, r_m \in R$, then $t_i \circ r_j = r_j \circ t_i$ for all i and j by Assumption 4, whence $t \circ r = t_n \circ \cdots \circ t_1 \circ r_m \circ \cdots \circ r_1 = r_m \circ \cdots \circ r_1 \circ t_n \circ \cdots \circ t_1 = r \circ t$. □

Lemma 3.3. *A reduction is uniquely determined by its values on $\mathcal{Y}$.*

Proof. Let $t \in T(S)$ be arbitrary. Since t is continuous and $\mathcal{M}$ is dense in $\overline{\mathcal{M}}$, the values on $\mathcal{M}$ uniquely determine t. By Lemma 2.3, any $a \in \mathcal{M}$ can be expressed as $a = \sum_{k=1}^{n} r_k(\mu_k)$ for some $n \in \mathbb{N}$, $r_1, \ldots, r_n \in \pm R^*$, and $\mu_1, \ldots, \mu_n \in \mathcal{Y}$. Hence $t(a) = \sum_{k=1}^{n} t\big(r_k(\mu_k)\big) = \sum_{k=1}^{n} r_k\big(t(\mu_k)\big)$, which expresses $t(a)$ purely in terms of the values on $\mathcal{Y}$ of t. □

The definitions of reductions are accordingly often simplified to stating how they act on elements of $\mathcal{Y}$. A common approach is to define simple reductions so that they change precisely one element of $\mathcal{Y}$ while leaving all other elements the same. Concretely the simple reduction which changes $\mu \in \mathcal{Y}$ to $a \in \overline{\mathcal{M}}$ would be defined by

$$t_{\mu \mapsto a}(\lambda) = \begin{cases} a & \text{if } \lambda = \mu, \\ \lambda & \text{otherwise,} \end{cases} \qquad \text{for all } \lambda \in \mathcal{Y}. \tag{3.1}$$

Such a map satisfies $t_{\mu \mapsto a}(b) = b$ for all $b \in \mathrm{Cspan}\big(\mathcal{Y} \setminus \{\mu\}\big)$, so if in addition $a \in \mathrm{Cspan}\big(\mathcal{Y} \setminus \{\mu\}\big)$ (which will be hard to avoid while satisfying the compatibility condition of Definition 5.3), then the image of $t_{\mu \mapsto a}$ will be contained in $\mathrm{Cspan}\big(\mathcal{Y} \setminus \{\mu\}\big)$ and consequently this map becomes a projection. Its kernel is however as small as it can possibly be without being trivial, and the a is only rarely zero, which means the projection is typically skew.

While (3.1) is the standard definition of a simple reduction from a conceptual point of view, it is not obviously one which is formally sound; if for example $\mathcal{Y}$ is not an independent set in $\mathcal{M}$ then a map $t_{\mu \mapsto a}$ for arbitrary $\mu \in \mathcal{Y}$ and $a \in \overline{\mathcal{M}}$ can probably not both be a group homomorphism and satisfy (3.1). An alternative definition of $t_{\mu \mapsto a}$, which often is better suited for proving properties of this reduction, is

$$t_{\mu \mapsto a}(b) = b - f_\mu(b) \cdot (\mu - a) \qquad \text{for all } b \in \overline{\mathcal{M}}. \tag{3.2}$$

Prerequisites for this formula is that $\overline{\mathcal{M}}$ is some sort of $\mathcal{R}$-module that furthermore comes with coefficient-of-μ homomorphisms $f_\mu \colon \overline{\mathcal{M}} \longrightarrow \mathcal{R}$ for all

$\mu \in \mathcal{Y}$; if these satisfy $f_\mu(\mu) = 1$ and $f_\mu(\nu) = 0$ for all $\nu \in \mathcal{Y} \setminus \{\mu\}$ then (3.1) becomes an immediate consequence of (3.2). Assuming the $\mathcal{R}$-module operations on $\overline{\mathcal{M}}$ are continuous, the continuity of $t_{\mu\mapsto a}$ is furthermore implied by the continuity of the coefficient function f_μ, and this depends only on the choices of $\mathcal{M}$, R, $\mathcal{Y}$, and $\mathcal{O}$. With ultranorms defined using (2.6), the continuity of these f_μ maps is typically something one gets for free [7, Ssec. 2.3.2]. See also Lemma 7.1.

Formulae like (3.2) can often be used to define the simple reductions even in cases where (3.1) leads to contradictions due to dependencies between elements of $\mathcal{Y}$. One example of this is the situation that $\mathcal{M}$ is a free $\mathcal{R}$-module with basis Y, but R is less than $\mathcal{R}$ and $\mathcal{Y}$ therefore has been chosen as the set of all multiples $r\mu$ for $r \in \mathcal{R}$ and $\mu \in Y$. If the range of μ in (3.2) is restricted to Y then this formula still makes perfect sense, but the result is of course rather a map $t_{\mu\mapsto a}$ satisfying

$$t_{\mu\mapsto a}(\lambda) = \begin{cases} ra & \text{if } \lambda = r\mu \text{ for some } r \in \mathcal{R}, \\ \lambda & \text{otherwise,} \end{cases} \qquad \text{for all } \lambda \in \mathcal{Y}.$$

The underlying idea for all these definitions of a reduction $t_{\mu\mapsto a}$ is to distinguish the part of a general element of $\overline{\mathcal{M}}$ that corresponds to the particular undesired element μ of $\mathcal{Y}$, and then replace this part by something it is equivalent to. This process is often straightforward for concrete problems, even though it may seem difficult to formalise in general.

Definition 3.4. A reduction $t \in T(S)$ is said to **act trivially** on some $a \in \overline{\mathcal{M}}$ if $t(a) = a$. An element $a \in \overline{\mathcal{M}}$ is said to be **irreducible** (with respect to $T(S)$) if all $t \in T(S)$ act trivially on it. The set of all irreducible elements in $\overline{\mathcal{M}}$ is denoted $\operatorname{Irr}(S)$. Also let

$$\mathcal{I}(S) = \overline{\sum_{t\in T(S)} \left\{ a - t(a) \,\middle|\, a \in \overline{\mathcal{M}} \right\}} \tag{3.3}$$

and write $a \equiv b \pmod{S}$ for $a - b \in \mathcal{I}(S)$. An $a \in \operatorname{Irr}(S)$ is said to be a **normal form of** $b \in \overline{\mathcal{M}}$ if $a \equiv b \pmod{S}$.

The main theme in the next two sections is to define a projection t^S of $\overline{\mathcal{M}}$ onto $\operatorname{Irr}(S)$ that constitutes a kind of pointwise limit of $T(S)$, and then demonstrate that $\mathcal{I}(S)$ is the kernel of that projection; from this will follow that there exists a unique normal form (which is computed by the map t^S) for every element of $\overline{\mathcal{M}}$. When it all works out, there is an equivalence

$$a \equiv b \pmod{S} \quad \Longleftrightarrow \quad t^S(a) = t^S(b) \qquad \text{for all } a, b \in \overline{\mathcal{M}}, \tag{3.4}$$

where the right hand side is algorithmic in style and well suited for calculations, whereas the congruence relation $\equiv$ in the left hand side is identifiable as the reflexive–symmetric–transitive–algebraic–topological closure of

'$a \equiv t(a)$ for all $t \in T_1(S)$ and $a \in \overline{\mathcal{M}}$'. A problem that is difficult on one side of the equivalence may have an obvious solution when transported to the other side of it; the main direction for decision problems is left to right, whereas identities tend to be simpler to derive on the left side. Applied calculations often focus on the irreducible elements, because the set $\mathrm{Irr}(S)$ can be used as a model for the quotient set $\overline{\mathcal{M}}/\!\equiv \pmod{S}$.

It should be pointed out that this concept of irreducibility has nothing to do with *multiplicative irreducibility* (the property that the only factorisations of an element are the trivial ones), nor for that matter with for example *join-irreducibility* (which in lattice theory is the equally important property that an element cannot be expressed as the $\vee$ of two other elements), but the point about multiplicative irreducibility needs to be stressed since many algebraists are accustomed to interpreting an unqualified 'irreducible' as referring to precisely multiplicative irreducibility; indeed this tendency is so strong that many authors seek other names to use for this concept. One such synonym is *normal* (which in this sense most commonly occurs in the phrase 'normal form'), that unfortunately also has the alternative interpretations "having norm 1" and "being orthogonal to tangents", which are quite different. Another synonym is *terminal*, which refers to the fact that reduction stops when reaching one of these elements — however in this topologized setting reductions do not in general stop completely; they merely "slow down" when approaching the limit. 'Irreducible' is the term that is used in [2] and must therefore be considered established and standard.

Lemma 3.5. *An element of $\overline{\mathcal{M}}$ is irreducible if and only if every simple reduction acts trivially on it. The set $\mathrm{Irr}(S)$ is a topologically closed R-module. More generally, any continuous group homomorphism $r\colon \overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ satisfying $r \circ t = t \circ r$ for all $t \in T(S)$ maps $\mathrm{Irr}(S)$ into itself.*

Proof. Clearly every simple reduction acts trivially on an irreducible element. Conversely every non-simple reduction is a composition of simple reductions, and if all of these act trivially on an element, then the composite reduction must do so too. Hence all elements which simple reductions act trivially upon are irreducible.

Let $t \in T(S)$ be arbitrary. Then the set I_t of all $b \in \overline{\mathcal{M}}$ such that $t(b) = b$ can alternatively be characterised as the kernel of the map $t'\colon \overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ defined by $t'(b) = t(b) - b$. Since t is a continuous group homomorphism, t' will be one too, and thus the set I_t will be a subgroup of $\overline{\mathcal{M}}$. Moreover I_t is topologically closed since it is the inverse image of $\{0\}$, which is a closed set. If r is a homomorphism commuting with t then for any $b \in I_t$, $0 = r\big(t'(b)\big) = (r \circ t)(b) - r(b) = (t \circ r)(b) - r(b) = t'\big(r(b)\big)$, and hence r maps I_t into itself. Finally $\mathrm{Irr}(S) = \bigcap_{t \in T(S)} I_t$ and hence $\mathrm{Irr}(S)$ must also be a topologically closed R-module, since these properties are preserved under arbitrary intersections. □

There is a similar set of basic properties that hold for the complementary set $\mathcal{I}(S)$ of elements equivalent to 0, but before going into that it is convenient to expound a bit on some additional twists in the usual construction of simple reductions. It was said above that S could be a set of "directed equivalences", usually known as *rewrite rules* or simply *rules*. Concretely these rules may be expressed as pairs $(\mu, a) \in \mathcal{Y} \times \overline{\mathcal{M}}$ where μ (the 'principal' or 'leading' part) is to be replaced by a. The homomorphism $t_{\mu\mapsto a}$ as constructed in (3.1) or (3.2) implements this replacement, but $T_1(S)$ typically contains more than just those $t_{\mu\mapsto a}$ maps for which $(\mu, a) \in S$; there will also be reductions which arise from placing the basic rule into various contexts. This produces reductions that apply in cases where μ occurs as a part of a larger expression.

In the classical case of Bergman's diamond lemma, where $\overline{\mathcal{M}} = \mathcal{R}\langle X\rangle$ and $\mathcal{Y} = X^*$, this means that a pair $(\mu, a) \in S$ should not only give rise to a simple reduction which maps μ to a, it should also for every multiple $\nu_1\mu\nu_2$ of μ give rise to a simple reduction designed to map that $\nu_1\mu\nu_2$ to the corresponding multiple $\nu_1 a\nu_2$ of a. Thus if one defines

$$t_{\nu_1 s\nu_2}(\lambda) = \begin{cases} \nu_1 a_s \nu_2 & \text{if } \lambda = \nu_1\mu_s\nu_2, \\ \lambda & \text{otherwise,} \end{cases} \tag{3.5}$$

for all $(\mu_s, a_s) = s \in S$ and $\lambda, \nu_1, \nu_2 \in \mathcal{Y}$, then the corresponding construction of $T_1(S)$ is

$$T_1(S) = \{\, t_{\nu_1 s\nu_2} \,\big|\, \nu_1, \nu_2 \in \mathcal{Y}, s \in S \,\}\,. \tag{3.6}$$

In this case, it follows that a monomial $\lambda \in X^*$ is irreducible if and only if it is not a multiple of μ_s for any $s \in S$. If S is finite then the irreducible words furthermore constitute a regular language (i.e., they can be described by a regexp), although infinite rewriting systems are unavoidable for some equivalences. Conversely, it would typically not be possible to make do with a finite system S unless there was this kind of "put rule into all possible contexts" mechanism manufacturing an infinite family of reductions from every concrete rule. Though this twist is not a technical necessity, it is in practice very helpful.

The basic idea of putting rules into all possible contexts remains useful in general, but beyond associative algebra it quickly becomes difficult to express concretely in elementary notation. The abstract form of this construction is that one has a set V of continuous homomorphisms $\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$, which furthermore map $\mathcal{Y}$ into itself, and defines simple reductions $t_{v,s}$ through

$$t_{v,s}(\lambda) = \begin{cases} v(a_s) & \text{if } \lambda = v(\mu_s), \\ \lambda & \text{otherwise,} \end{cases} \tag{3.7}$$

for all $(\mu_s, a_s) = s \in S$ and $v \in V$. The set of maps which gives rise to the classical case described above is

$$V = \{b \mapsto \nu_1 b \nu_2\}_{\nu_1,\nu_2 \in X^*}, \tag{3.8}$$

but one can also consider other families — see below for some examples.

If V (as above) is a monoid under composition then the construction (3.7) has the effect that there for every such reduction $t_{v,s}$, every $w \in V$, and every $\lambda \in \mathcal{Y}$ exists another reduction $t_{w\circ v,s}$ which satisfies $t_{w\circ v,s}\big(w(\lambda)\big) = w\big(t_{v,s}(\lambda)\big)$, and if w is injective then it's even $t_{w\circ v,s} \circ w = w \circ t_{v,s}$. This turns out to be a very useful property, so it deserves a name.

Definition 3.6. A map $v\colon \overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ is said to be **advanceable** (with respect to $T_1(S)$) if there for every $t \in T_1(S)$ and $b \in R^*\mathcal{Y}$ exists some $u \in T(S)$ such that $u\big(v(b)\big) = v\big(t(b)\big)$. The map v is said to be **absolutely advanceable** (with respect to $T(S)$) if there for every $t \in T(S)$ exists some $u \in T(S)$ such that $v \circ t = u \circ v$. For contrast, the ordinary advanceability may also be called *conditional advanceability*.

Advanceable maps provide a way to reason about and take advantage of the kind of structures in the set of reductions that arise from using (3.7) or some variation thereof. The name comes from the point of view that if an advanceable map v and a reduction are both to be applied to some element, then one can always arrange things so that v is applied *before* the reduction (it can be *advanced* past a reduction), even if that may come at the price of having to apply a different reduction. A point worth noticing is that absolute advanceability only needs to be checked for simple reductions, as a map v can be advanced past $t_1 \circ t_2$ if it can be advanced past t_1 and t_2, whereas a conditional advancement need not have this composition property; different terms of $t_2(b)$ may call for different translations of t_1 when one tries to advance v past t_1. It is however the conditional variant that in practice is most important.

Absolute advanceability can be viewed as a weaker form of the 'more generally' condition in Lemma 3.5, in that it doesn't require the two reductions to be equal; this point of view is employed in the next lemma. Another way of expressing this condition is that $v \circ T(S) \subseteq T(S) \circ v$, and therefore some may prefer to describe an absolute advanceable map as being an element of the left-normaliser of $T(S)$, but that characterisation seems difficult to use for conditional advanceability. Furthermore the characterisation as element of the left-normaliser breaks down in the more general case of a multi-sorted structure; Definition 6.1 gives the whole story.

Lemma 3.7. *The set $\mathcal{I}(S)$ is a topologically closed R-module. More generally, any advanceable continuous group endomorphism on $\overline{\mathcal{M}}$ maps $\mathcal{I}(S)$ into itself. Any reduction maps $\mathcal{I}(S)$ into itself, and in particular* $\ker t \subseteq \mathcal{I}(S)$

for all $t \in T(S)$. Furthermore

$$\mathcal{I}(S) = \overline{\sum_{t\in T_1(S)} \{\, a - t(a) \,\big|\, a \in \overline{\mathcal{M}} \,\}} = \overline{\sum_{t\in T_1(S)} \{\, a - t(a) \,\big|\, a \in \mathcal{M} \,\}} = \\ = \operatorname{Cspan}\Big(\{\, \mu - t(\mu) \,\big|\, \mu \in \mathcal{Y}, t \in T_1(S) \,\}\Big). \quad (3.9)$$

Proof. Let $t \in T(S)$ be arbitrary. For any $b, c \in \overline{\mathcal{M}}$ one finds that

$$\big(b{-}t(b)\big){-}\big(c{-}t(c)\big) = b{-}c{-}t(b){+}t(c) = (b{-}c){-}t(b{-}c) \in \{\, a - t(a) \,\big|\, a \in \overline{\mathcal{M}} \,\}$$

and hence $N_t := \{\, a - t(a) \,\big|\, a \in \overline{\mathcal{M}} \,\}$ is a subgroup of $\overline{\mathcal{M}}$. Obviously the sum of a family of groups is a group, and the closure of a group is a group because the group operation (addition) is continuous. Hence $\mathcal{I}(S)$ is a group, and it is topologically closed by definition. From the similar observation that any $a \in \ker t$ satisfies $a = a - t(a) \in N_t \subseteq \mathcal{I}(S)$, it follows that $\ker t \subseteq \mathcal{I}(S)$.

Now let $r\colon \overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ be an absolutely advanceable continuous group homomorphism. Let $t \in T(S)$ be arbitrary and choose some $t' \in T(S)$ such that $r \circ t = t' \circ r$. For any $b \in \overline{\mathcal{M}}$ one finds that

$$r\big(b - t(b)\big) = r(b) - r\big(t(b)\big) = r(b) - t'\big(r(b)\big) \in N_{t'}$$

and hence r maps N_t into $N_{t'}$. It follows from the fact that r is a homomorphism that r maps $N = \sum_{t\in T(S)} N_t$ into itself, and then from the fact that r is continuous that it maps $\mathcal{I}(S) = \overline{N}$ into itself. Since in particular all $r \in R$ are absolutely advanceable, it follows that $\mathcal{I}(S)$ is an R-module.

Again let $t \in T(S)$ be arbitrary and consider the matter of whether t maps $\mathcal{I}(S)$ into itself. For any $t' \in T(S)$ and $b \in N_{t'}$ it holds that $b = a{-}t'(a)$ for some $a \in \overline{\mathcal{M}}$, and thus

$$t(b) = t(a) - (t \circ t')(a) = a - (t \circ t')(a) - a + t(a) \in N_{t\circ t'} + N_t \subseteq \mathcal{I}(S).$$

Hence $t(N_{t'}) \subseteq \mathcal{I}(S)$ for all $t' \subseteq T(S)$ and this extends as above to arbitrary elements of $\mathcal{I}(S)$; an arbitrary reduction $t \in T(S)$ maps $\mathcal{I}(S)$ into itself.

For the claim that $\mathcal{I}(S)$ can be constructed from the N_t groups of simple reductions, one may first observe that N_{id} is just $\{0\}$, and thus does not contribute anything unique to $\mathcal{I}(S)$. Any other nonsimple reduction $t \in T(S)$ is a finite composition $t_n \circ \cdots \circ t_1 = t$ of simple reductions $t_1, \ldots, t_n \in T(S)$ and it holds that $N_t \subseteq \sum_{k=1}^{n} N_{t_k}$, because if $u_k = t_k \circ \cdots \circ t_1$ for $k = 1, \ldots, n$ then any $a{-}t(a) \in N_t$ can be written as $a{-}t_1(a){+}u_1(a){-}t_2\big(u_1(a)\big){+}\cdots + u_{n-1}(a) - t_n\big(u_{n-1}(a)\big) \in N_{t_1} + N_{t_2} + \cdots + N_{t_n}$. Hence $\sum_{t\in T(S)} N_t \subseteq \sum_{t\in T_1(S)} N_t$; the terms for simple reductions suffice for producing the total sum.

Define $M_t := \{\, a - t(a) \,\big|\, a \in \mathcal{M} \,\}$ for all $t \in T_1(S)$. The last equality in (3.9) follows from the observation that $M_t = \operatorname{Span}\Big(\{\, \mu - t(\mu) \,\big|\, \mu \in \mathcal{Y} \,\}\Big)$

for all $t \in T_1(S)$. In the middle equality, the $\supseteq$ inclusion trivially follows from $N_t \supseteq M_t$. For the reverse inclusion, let b in the closure of $\sum_{t\in T_1(S)} N_t$ be given. Let $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary. Since $b+\varepsilon$ is a neighbourhood of b it contains some element of $\sum_{t\in T_1(S)} N_t$, i.e., there exists a finite $U \subseteq T_1(S)$ and $\{a_t\}_{t\in U} \subseteq \overline{\mathcal{M}}$ such that $b - \sum_{t\in U}\big(a_t - t(a_t)\big) \in \varepsilon$. Let $\delta \in \widehat{\mathcal{O}}$ be such that $\delta \subseteq \varepsilon$ and $t(\delta) \subseteq \varepsilon$ for all $t \in U$. Let $\{c_t\}_{t\in U} \subseteq \mathcal{M}$ be such that $a_t - c_t \in \delta$ for all $t \in U$. Then

$$\sum_{t\in U}\big(c_t - t(c_t)\big) = \sum_{t\in U}\big(a_t - t(a_t)\big) + \sum_{t\in U}(c_t - a_t) - \sum_{t\in U} t(c_t - a_t) \in b+\varepsilon+\delta-\varepsilon = b+\varepsilon$$

and hence b, by the arbitrariness of ε, is in the closure of $\sum_{t\in T_1(S)} M_t$.

The claim that also a conditionally advanceable continuous homomorphism v will map $\mathcal{I}(S)$ into itself is now an easy consequence of (3.9): by continuity and since v is a homomorphism, it suffices to show $v(b) \in \mathcal{I}(S)$ for arbitrary $b \in \mathrm{Span}\big(\{\mu - t(\mu)\}\big)$, $\mu \in \mathcal{Y}$, and $t \in T_1(S)$. Let such b, μ, and t be given. There exist $\{r_i\}_{i=1}^n \subseteq \pm R^*$ such that $b = \sum_{i=1}^n r_i\big(\mu - t(\mu)\big)$ and $\{t_i\}_{i=1}^n \in T(S)$ such that $(t_i \circ v \circ r_i)(\mu) = (v \circ t \circ r_i)(\mu)$. Thus

$$\begin{aligned} v(b) = v\bigg(\sum_{i=1}^n r_i\big(\mu - t(\mu)\big)\bigg) &= \sum_{i=1}^n \bigg(v\big(r_i(\mu)\big) - v\Big(r_i\big(t(\mu)\big)\Big)\bigg) = \\ &= \sum_{i=1}^n \bigg(v\big(r_i(\mu)\big) - t_i\Big(v\big(r_i(\mu)\big)\Big)\bigg) \in \mathcal{I}(S). \end{aligned}$$

□

In the case of Bergman's diamond lemma, where all maps on the form (3.8) are advanceable, this lemma implies that $\mathcal{I}(S)$ is a two-sided ideal: it is closed under addition, multiplication by a scalar, and multiplication on either side by an arbitrary generator of the algebra $\overline{\mathcal{M}} = \mathcal{R}\langle X\rangle$, so by distributivity it is closed under multiplication by arbitrary elements. The letter '$\mathcal{I}$' was chosen in anticipation of this, since 'I' is the initial of 'ideal', but it is by no means restricted to two-sided ideals.

Definition 3.8. Let V be a set of maps $\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$. A nonempty R-module $N \subseteq \overline{\mathcal{M}}$ is said to be a **V-ideal** if it is topologically closed and $v(a) \in N$ for all $v \in V$ and $a \in N$. A set $A \subseteq \overline{\mathcal{M}}$ is said to be a **V-ideal basis** for N if $N = \mathrm{Cspan}\big(\{v(a)\}_{a\in A, v\in V}\big)$.

From a minimalistic formal perspective the 'V-ideal' concept is unnecessary, as it is equivalent to 'topologically closed $(R\cup V)$-module', but (apart from being shorter) the 'V-ideal' terminology has the advantage of being closer to the familiar terms 'left ideal', 'right ideal', and 'two-sided ideal' of which 'V-ideal' is a common generalisation. Furthermore, it is 'V-ideal basis' that is the more important concept in the above definition, since that

is a first step on the way to defining a *Gröbner basis*. Ideal bases have the right associations for this, whereas any combination of 'module' and 'basis' is likely to give rise to incorrect expectations about independence between basis elements.

Besides preparing for subsequent developments, this definition also gives an opportunity to summarise some of the above constructions of simple reductions into a formal statement that actually claims something, even if it isn't very much.

Corollary 3.9. *Let $\mathcal{R} \supseteq R$ be a unital ring of continuous endomorphisms on $\overline{\mathcal{M}}$ that is equipped with a topology such that the $\mathcal{R}$-module action $\mathcal{R}\times\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}} : (r, b) \mapsto r(b) =: r \cdot b$ is continuous and $\mathcal{R}$ is complete. Assume $\mathcal{M}$ is a free $\mathcal{R}$-module with basis $\mathcal{Y}$ such that each coefficient-of-μ homomorphism $f_\mu\colon \mathcal{M} \longrightarrow \mathcal{R}$ is continuous.*

If V is a monoid of continuous $\mathcal{R}$-module homomorphisms $\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ that map $\mathcal{Y}$ into itself then the following holds for all $S \subseteq \mathcal{Y} \times \overline{\mathcal{M}}$:

1. *For any $v \in V$ and $s = (\mu_s, a_s) \in S$, the map $t_{v,s}\colon \overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ defined by*

$$t_{v,s}(b) = b - f_{v(\mu_s)}(b)\big(v(\mu_s) - v(a_s)\big) \qquad \text{for all } b \in \overline{\mathcal{M}} \tag{3.10}$$

   is a continuous homomorphism that commutes with all elements of $\mathcal{R}$.

2. *If $T_1(S) = \{t_{v,s}\}_{v\in V, s\in S}$ then every $v \in V$ is advanceable and the set $\{\mu_s - a_s\}_{s\in S}$ is a V-ideal basis for $\mathcal{I}(S)$.*

3. *Any injective element of V is absolutely advanceable.*

Proof. The definition (3.10) of $t_{v,s}$ is clearly a composition of maps that by assumption are continuous homomorphisms. Furthermore $t_{v,s}(r \cdot b) = r \cdot b - f_{v(\mu_s)}(r \cdot b) \cdot v(\mu_s - a_s) = r \cdot b - \big(r \circ f_{v(\mu_s)}(b)\big) \cdot v(\mu_s - a_s) = r \cdot t_{v,s}(b)$ for all $r \in \mathcal{R}$ and $b \in \overline{\mathcal{M}}$ since $f_{v(\mu_s)}$ is an $\mathcal{R}$-module homomorphism and $\cdot$ is a left module action. Hence $t_{v,s} \circ r = r \circ t_{v,s}$.

By the last part of (3.9), $\mathcal{I}(S)$ is spanned by all $\lambda - t(\lambda)$ for $\lambda \in \mathcal{Y}$ and $t \in T_1(S)$, i.e., all $\lambda - t_{v,s}(\lambda) = \lambda - \lambda + f_{v(\mu_s)}(\lambda) \cdot v(\mu_s - a_s)$. This is 0 unless $\lambda = v(\mu_s)$, in which case it is equal to $v(\mu_s - a_s)$. Hence $\mathrm{Cspan}\big(\{v(\mu_s - a_s)\}_{v\in V, s\in S}\big) = \mathcal{I}(S)$.

In order to show that $w \in V$ is advanceable, let $t_{v,s} \in T_1(S)$ and $\lambda \in \mathcal{Y}$ be arbitrary. If $\lambda = v(\mu_s)$ then

$$\begin{aligned} w\big(t_{v,s}(\lambda)\big) = w\Big(\lambda - f_{v(\mu_s)}(\lambda) \cdot v(\mu_s - a_s)\Big) = w\big(v(\mu_s) - v(\mu_s - a_s)\big) = \\ = (w \circ v)(a_s) = (w \circ v)(\mu_s) - (w \circ v)(\mu_s - a_s) = \\ = w(\lambda) - f_{(w\circ v)(\mu_s)}\big((w \circ v)(\mu_s)\big) \cdot (w \circ v)(\mu_s - a_s) = t_{w\circ v,s}\big(w(\lambda)\big) \end{aligned}$$

and since w, $t_{v,s}$, and $t_{w\circ v,s}$ are all $\mathcal{R}$-module homomorphisms it follows that $w\big(t_{v,s}(r\cdot\lambda)\big) = t_{w\circ v,s}\big(w(r\cdot\lambda)\big)$ for all $r\in\mathcal{R}$. If instead $\lambda \neq v(\mu_s)$ then

$$w\big(t_{v,s}(\lambda)\big) = w\Big(\lambda - f_{v(\mu_s)}(\lambda)\cdot v(\mu_s - a_s)\Big) = w(\lambda) = \mathrm{id}\big(w(\lambda)\big)$$

and since w, $t_{v,s}$, and id are all $\mathcal{R}$-module homomorphisms it follows that $w\big(t_{v,s}(r\cdot\lambda)\big) = \mathrm{id}\big(w(r\cdot\lambda)\big)$ for all $r\in\mathcal{R}$. Either way, w can be advanced past $t_{v,s}$ and hence w is advanceable.

If w is injective then $w(\lambda) = (w\circ v)(\mu_s)$ if and only if $\lambda = v(\mu_s)$ and hence $f_{(w\circ v)(\mu_s)}\circ w = f_{v(\mu_s)}$. In this case $w\big(t_{v,s}(b)\big) = t_{w\circ v,s}\big(w(b)\big)$ for all $b\in\overline{\mathcal{M}}$. □

In the case $\mathcal{M} = \mathcal{R}\langle X\rangle$, the choices of V that give rise to one-sided ideals are

$$V = \{b\mapsto \nu b\}_{\nu\in X^*} \qquad \text{(left ideal)},$$
$$V = \{b\mapsto b\nu\}_{\nu\in X^*} \qquad \text{(right ideal)};$$

multiplication by a non-monomial element of $\mathcal{M}$ does not map $\mathcal{Y} = X^*$ into itself and can therefore not be used with Corollary 3.9. In the fourth classical case that $\mathcal{M}$ is an $\mathcal{R}[X]$-module and one wants $\mathcal{I}(S)$ to be an $\mathcal{R}[X]$-submodule, the right choice is to make V the set of all maps $b\mapsto b\prod_{x\in X} x^{n_x}$ for $\{n_x\}_{x\in X}\subset\mathbb{N}$, i.e., the set of maps that multiply by power products on X.

4 The limit of all reductions

The head-on approach for defining the sought projection t^S of $\overline{\mathcal{M}}$ onto $\mathrm{Irr}(S)$ would be to immediately seek a map $\overline{\mathcal{M}}\longrightarrow\mathrm{Irr}(S)$, but often a subtler approach is more convenient. The route taken here is to (i) define subsets of $\overline{\mathcal{M}}$ where the collective of reductions has nice properties, (ii) use those properties in the definition of t^S, and only afterwards (iii) show that the subsets defined in (i) are in fact the whole of $\overline{\mathcal{M}}$; the same approach was followed in [2]. Steps (i) and (ii) are carried out in this section, whereas step (iii) is the subject of the next.

The definition of t^S as a pointwise limit of $T(S)$ would be to say that $t^S(a)$ is the element of $\mathrm{Irr}(S)$ that is a limit point of $\big\{\, t(a)\;\big|\; t\in T(S)\,\big\}$; in a discrete topology, this simply says that $t^S(a)$ is the element of $\mathrm{Irr}(S)$ which is equal to $t(a)$ for some $t\in T(S)$, but in general one will have to make do with being able to get arbitrarily close to some irreducible element. For this idea to work as a definition of $t^S(a)$ it is of course first necessary that such a limit point exists (and second necessary that it is unique), but the existence condition becomes much more convenient if one strengthens it a bit.

Definition 4.1. An $a \in \overline{\mathcal{M}}$ is said to be **stuck in** $N \subseteq \overline{\mathcal{M}}$ under $T(S)$ if $t(a) \in N$ for all $t \in T(S)$. Given an $\varepsilon \in \widehat{\mathcal{O}}$, an $a \in \overline{\mathcal{M}}$ is said to be **persistently ε-reducible** under $T(S)$ if there for every $t_1 \in T(S)$ exists some $t_2 \in T(S)$ and $b \in \mathrm{Irr}(S)$ such that $t_2\big(t_1(a)\big)$ is stuck in $b + \varepsilon$.

If a is persistently ε-reducible for all $\varepsilon \in \widehat{\mathcal{O}}$ then a is said to be **persistently reducible**. The set of all elements in $\overline{\mathcal{M}}$ that are persistently reducible under $T(S)$ is denoted $\mathrm{Per}(S)$ and the set of all elements in $\overline{\mathcal{M}}$ that are persistently ε-reducible under $T(S)$ is denoted $\mathrm{Per}_\varepsilon(S)$.

The extra power added in this definition is that the wanted outcome of being close to $\mathrm{Irr}(S)$ should *persist* no matter what has to be done before or after the t_2 reduction chosen to get there: $a \in \mathrm{Per}_\varepsilon(S)$ if and only if there for every $t_1 \in T(S)$ exists some $t_2 \in T(S)$ and $b \in \mathrm{Irr}(S)$ such that it holds for every $t_3 \in T(S)$ that $(t_3 \circ t_2 \circ t_1)(a) - b \in \varepsilon$. The corresponding property in [2] is 'reduction-finiteness'; see [7, Ssec. 3.1.2] for a comparison and analysis of the two.

Lemma 4.2. *For each $a \in \mathrm{Per}(S)$, there exists some $b \in \mathrm{Irr}(S)$ such that for every $\varepsilon \in \widehat{\mathcal{O}}$ there is some $t \in T(S)$ such that $t(a)$ is stuck in $b + \varepsilon$. Furthermore a and b are such that $a - b \in \mathcal{I}(S)$, and hence* $\mathrm{Per}(S) \subseteq \mathcal{I}(S) + \mathrm{Irr}(S)$.

Proof. Construct a sequence $\{u_i\}_{i=0}^{\infty} \subseteq T(S)$ by letting $u_0 = \mathrm{id}$ and recursively defining u_n for $n > 0$ as follows: by persistent reducibility of a there exist $t_n \in T(S)$ and $b_n \in \mathrm{Irr}(S)$ such that $t_n\big(u_{n-1}(a)\big)$ is stuck in $b_n + \overline{B_n}$, therefore let $u_n = t_n \circ u_{n-1}$.

For any $n, i, j \in \mathbb{N}$ such that $i > j \geqslant n$, the element $u_i(a)$ is in $b_i + \overline{B_i}$ as well as in $b_j + \overline{B_j}$, which means

$$b_i - b_j = \big(b_i - u_i(a)\big) + \big(u_i(a) - b_j\big) \in \overline{B_i} + \overline{B_j} \subseteq \overline{B_n}.$$

Hence $\{b_n\}_{n=0}^{\infty}$ is a Cauchy sequence in $\mathrm{Irr}(S)$ and consequently it converges to some $b \in \mathrm{Irr}(S)$. Fix some $n \in \mathbb{N}$ such that $\overline{B_n} \subseteq \varepsilon$ and consider $\lim_{i\to\infty}(b_i - b_n)$. Since all elements in this sequence are in the closed set $\overline{B_n}$ it follows that the limit $b - b_n \in \overline{B_n} \subseteq \varepsilon$. Thus $b_n + \varepsilon = b + \varepsilon$ and $u_n(a)$ is stuck in $b + \varepsilon$ as claimed.

To see the second claim, observe that $u_n(a) \to b$ as $n \to \infty$. For any $n \in \mathbb{N}$, $a - u_n(a) \in \mathcal{I}(S)$ by definition and consequently $a - b = \lim_{n\to\infty}\big(a - u_n(a)\big) \in \mathcal{I}(S)$ as well. Hence $a \in \mathcal{I}(S) + \mathrm{Irr}(S)$. □

This proof highlights a subtle point in the basic set-up of Section 2 which may be regarded as a restriction, namely that the family $\mathcal{O} = \{B_n\}_{n=1}^{\infty}$ must be countable: the construction of $\{u_n\}_{n=0}^{\infty}$ would not necessarily suffice for demonstrating convergence if $\mathcal{O}$ was uncountable. In other proofs it is possible to treat $\widehat{\mathcal{O}}$ as an arbitrary collection of neighbourhoods, which might suggest a generalisation to topologies defined by an uncountable $\mathcal{O}$ is not

unreasonable, but on the other hand the countability of $\mathcal{O}$ is in this lemma closely tied to the status of $T(S)$ as a set of finite compositions of elements of $T_1(S)$, and *that* is something several proofs rely on. Hence removing the condition that $\mathcal{O}$ is countable would probably require a more powerful construction of reductions than as mere compositions of simple reductions; it is presently not something for which I see any precedence.

Besides this existence of a limit point property, it is also important that the set of persistently reducible elements is closed under algebraic operations. This exercises the slightly different aspect of persistent reducibility that one can find reductions which simultaneously take several persistently reducible elements close to their normal forms.

Lemma 4.3. *The set* $\mathrm{Per}(S)$ *and for every* $\varepsilon \in \widehat{\mathcal{O}}$ *the set* $\mathrm{Per}_\varepsilon(S)$ *are* R*-modules. More generally,* $\mathrm{Per}(S)$ *is mapped into itself by every continuous group homomorphism* $\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ *which commutes with all reductions.*

Proof. Let $\varepsilon \in \widehat{\mathcal{O}}$ be given. Clearly $\mathrm{Irr}(S) \subseteq \mathrm{Per}_\varepsilon(S)$, and hence to see that the latter is a group, it suffices to show for two arbitrary elements in it that their difference is also in this set. Therefore let $a_1, a_2 \in \mathrm{Per}_\varepsilon(S)$ and $t_1 \in T(S)$ be arbitrary. There exists some $t_2 \in T(S)$ and $b_1 \in \mathrm{Irr}(S)$ such that $t_2\big(t_1(a_1)\big)$ is stuck in $b_1 + \varepsilon$. There also exists some $t_3 \in T(S)$ and $b_2 \in \mathrm{Irr}(S)$ such that $t_3\big((t_2 \circ t_1)(a_2)\big)$ is stuck in $b_2 + \varepsilon$. Now let $t_4 \in T(S)$ be arbitrary. Since

$$\begin{aligned} &t_4\Big((t_3 \circ t_2)\big(t_1(a_1 - a_2)\big)\Big) - (b_1 - b_2) = \\ &= \Big((t_4 \circ t_3)\Big(t_2\big(t_1(a_1)\big)\Big) - b_1\Big) - \Big(t_4\Big(t_3\big((t_2 \circ t_1)(a_2)\big)\Big) - b_2\Big) \in \varepsilon + \varepsilon = \varepsilon, \end{aligned}$$

it follows that $t_4\Big((t_3 \circ t_2)\big(t_1(a_1 - a_2)\big)\Big) \in (b_1 - b_2) + \varepsilon$. Hence the element $(t_3 \circ t_2)\big(t_1(a_1 - a_2)\big)$ is stuck in $(b_1 - b_2) + \varepsilon$, and by arbitrariness of t_1 it follows that $a_1 - a_2 \in \mathrm{Per}_\varepsilon(S)$.

Now let $r\colon \overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ be an arbitrary continuous group homomorphism which satisfies $r \circ t = t \circ r$ for all $t \in T(S)$, and let $\delta \in \widehat{\mathcal{O}}$ be such that $r(\delta) \subseteq \varepsilon$. Let $a \in \mathrm{Per}_\delta(S)$ and $t_1 \in T(S)$ be arbitrary. Let $t_2 \in T(S)$ and $b \in \mathrm{Irr}(S)$ be such that $(t_2 \circ t_1)(a)$ is stuck in $b + \varepsilon$. Then for any $t_3 \in T(S)$,

$$\begin{aligned} (t_3 \circ t_2 \circ t_1)\big(r(a)\big) - r(b) = r\big((t_3 \circ t_2 \circ t_1)(a)\big) - r(b) = \\ = r\big((t_3 \circ t_2 \circ t_1)(a) - b\big) \in r(\delta) \subseteq \varepsilon \end{aligned}$$

and thus $(t_2 \circ t_1)\big(r(a)\big)$ is stuck in $r(b) + \varepsilon$. It follows that $r(a)$ is persistently ε-reducible.

In the case that $r \in R$, one knows from the fact that ε is an R-module that one can take $\delta = \varepsilon$, and then the above has shown that $\mathrm{Per}_\varepsilon(S)$ is an R-module. For more general r this need not be the case, and then one has

only shown about r that there for every $\varepsilon \in \widehat{\mathcal{O}}$ is some $\delta \in \widehat{\mathcal{O}}$ such that r maps $\mathrm{Per}_\delta(S)$ into $\mathrm{Per}_\varepsilon(S)$. Suppose now that $a \in \mathrm{Per}(S)$ is arbitrary, and consider the question of whether $r(a) \in \mathrm{Per}(S)$. Let $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary and let $\delta \in \widehat{\mathcal{O}}$ be such that $r\big(\mathrm{Per}_\delta(S)\big) \subseteq \mathrm{Per}_\varepsilon(S)$. Since $a \in \mathrm{Per}(S) \subseteq \mathrm{Per}_\delta(S)$, it follows that $r(a) \in \mathrm{Per}_\varepsilon(S)$, and hence $r(a) \in \bigcap_{\varepsilon \in \widehat{\mathcal{O}}} \mathrm{Per}_\varepsilon(S) = \mathrm{Per}(S)$ by the arbitrariness of ε. □

Definition 4.4. Let $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary. An $a \in \overline{\mathcal{M}}$ is said to be **ε-uniquely reducible** under $T(S)$ if, for any $t_1, t_2 \in T(S)$ and $b_1, b_2 \in \mathrm{Irr}(S)$ such that $t_1(a)$ is stuck in $b_1 + \varepsilon$ and $t_2(a)$ is stuck in $b_2 + \varepsilon$, it holds that $b_1 + \varepsilon = b_2 + \varepsilon$. The set of all elements in $\overline{\mathcal{M}}$ which are both persistently and ε-uniquely reducible under $T(S)$ is denoted $\mathrm{Red}_\varepsilon(S)$.

Lemma 4.5. *For every $\varepsilon \in \widehat{\mathcal{O}}$, the set $\mathrm{Red}_\varepsilon(S)$ is an R-module that furthermore is mapped into itself by every reduction.*

Proof. Let $\varepsilon \in \widehat{\mathcal{O}}$ be given. Let $a_1, a_2 \in \mathrm{Red}_\varepsilon(S)$ be arbitrary. It follows from Lemma 4.3 that $a_1 - a_2 \in \mathrm{Per}(S)$ and hence there exists $t_1 \in T(S)$ and $b \in \mathrm{Irr}(S)$ such that $t_1(a_1 - a_2)$ is stuck in $b + \varepsilon$. Since $a_1, a_2 \in \mathrm{Per}(S)$ there furthermore exist $t_2, t_3 \in T(S)$ and $b_1, b_2 \in \mathrm{Irr}(S)$ such that $t_2\big(t_1(a_1)\big)$ is stuck in $b_1 + \varepsilon$ and $t_3\big((t_2 \circ t_1)(a_2)\big)$ is stuck in $b_2 + \varepsilon$. This implies that, for $t = t_3 \circ t_2 \circ t_1$,

$$b - (b_1 - b_2) = b - t(a_1 - a_2) - \big(b_1 - t(a_1)\big) + \big(b_2 - t(a_2)\big) \in \varepsilon - \varepsilon + \varepsilon = \varepsilon.$$

Starting from some other $t_1' \in T(S)$ and $b' \in \mathrm{Irr}(S)$ such that $t_1'(a_1 - a_2)$ is stuck in $b' + \varepsilon$, one similarly gets the existence of $t_2', t_3' \in T(S)$ and $b_1', b_2' \in \mathrm{Irr}(S)$ such that $t_2'\big(t_1'(a_1)\big)$ is stuck in $b_1' + \varepsilon$ and $t_3'\big((t_2' \circ t_1')(a_2)\big)$ is stuck in $b_2' + \varepsilon$; in precisely the same way one furthermore shows that $b' - (b_1' - b_2') \in \varepsilon$. The ε-unique reducibility of a_1 and a_2 does however imply that $b_1 - b_1' \in \varepsilon$ and $b_2 - b_2' \in \varepsilon$. From this follows that $b - b' \in \varepsilon$ and hence $a_1 - a_2 \in \mathrm{Red}_\varepsilon(S)$. This has shown that $\mathrm{Red}_\varepsilon(S)$ is a group.

It must also be shown that elements of R map $\mathrm{Red}_\varepsilon(S)$ into itself. Let $r \in R$, $\varepsilon \in \widehat{\mathcal{O}}$, and $a \in \mathrm{Red}_\varepsilon(S)$ be given. Let $t_1, t_2 \in T(S)$ and $b_1, b_2 \in \mathrm{Irr}(S)$ be arbitrary such that $t_i\big(r(a)\big)$ is stuck in $b_i + \varepsilon$ for $i = 1, 2$. Since $a \in \mathrm{Per}(S)$, there exist $t_1', t_2' \in T(S)$ and $b_1', b_2' \in \mathrm{Irr}(S)$ such that $t_i'\big(t_i(a)\big)$ is stuck in $b_i' + \varepsilon$ for $i = 1, 2$. It follows that

$$b_i - r(b_i') = \big(b_i - (t_i' \circ t_i \circ r)(a)\big) + r\big((t_i' \circ t_i)(a) - b_i'\big) \in \varepsilon + r(\varepsilon) = \varepsilon$$

for $i = 1, 2$. Hence

$$b_1 - b_2 = b_1 - r(b_1') + r(b_1' - b_2') + r(b_2') - b_2 \in \varepsilon + r(\varepsilon) + \varepsilon = \varepsilon$$

and thus $r(a)$ is ε-uniquely reducible by the arbitrariness of t_1 and t_2. By Lemma 4.3, $r(a)$ is also persistently reducible, and so $r(a) \in \mathrm{Red}_\varepsilon(S)$.

The corresponding property for reductions is more trivial. ε-unique reducibility of $t(a)$ for $a \in \mathrm{Red}_\varepsilon(S)$ and $t \in T(S)$ is the claim that any $t_1, t_2 \in T(S)$ and $b_1, b_2 \in \mathrm{Irr}(S)$ such that $t_1\big(t(a)\big)$ is stuck in $b_1 + \varepsilon$ and $t_2\big(t(a)\big)$ is stuck in $b_2 + \varepsilon$ satisfy $b_1 + \varepsilon = b_2 + \varepsilon$, but that is just a special case of the ε-unique reducibility of a. □

Definition 4.6. An element in $\overline{\mathcal{M}}$ is said to be **uniquely reducible** if it is ε-uniquely reducible for all $\varepsilon \in \widehat{\mathcal{O}}$. The set of those element which are both persistently and uniquely reducible under $T(S)$ is denoted $\mathrm{Red}(S)$. Define the map $t^S \colon \mathrm{Red}(S) \longrightarrow \mathrm{Irr}(S)$ by letting $t^S(a)$ be the unique element of $\mathrm{Irr}(S)$ with the property that there for every $\varepsilon \in \widehat{\mathcal{O}}$ exists some $t \in T(S)$ such that $t(a)$ is stuck in $t^S(a) + \varepsilon$.

The chain of sets defined in this section is thus that $\mathrm{Red}(S) \subseteq \mathrm{Red}_\varepsilon(S) \subseteq \mathrm{Per}(S) \subseteq \mathrm{Per}_\varepsilon(S) \subseteq \overline{\mathcal{M}}$, with $\mathrm{Red}(S) = \bigcap_{\varepsilon \in \widehat{\mathcal{O}}} \mathrm{Red}_\varepsilon(S)$ and $\mathrm{Per}(S) = \bigcap_{\varepsilon \in \widehat{\mathcal{O}}} \mathrm{Per}_\varepsilon(S)$. The end one wants to see is that all of these are equal, and in Lemmas 5.5 and 5.8 this is taken care of by giving sufficient conditions for $\mathrm{Per}_\varepsilon(S) = \overline{\mathcal{M}}$ and $\mathrm{Red}_\varepsilon(S) = \overline{\mathcal{M}}$ respectively. A more immediate goal is however to establish that circumstances inside $\mathrm{Red}(S)$ are good.

Lemma 4.7. *The set* $\mathrm{Red}(S)$ *is an* R*-module that is mapped into itself by every reduction. The map* $t^S \colon \mathrm{Red}(S) \longrightarrow \mathrm{Irr}(S)$ *is well-defined and a group homomorphism. More generally, every continuous group homomorphism* $r \colon \overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ *which commutes with all reductions maps* $\mathrm{Red}(S)$ *into itself and commutes with* t^S*. In addition,* $t^S(b) = b$ *for all* $b \in \mathrm{Irr}(S)$*,* $\ker t^S \subseteq \mathcal{I}(S)$*, and* $t^S\big(t(a)\big) = t^S(a)$ *for all* $a \in \mathrm{Red}(S)$ *and* $t \in T(S)$*.*

Proof. Since $\mathrm{Red}(S) = \bigcap_{\varepsilon \in \widehat{\mathcal{O}}} \mathrm{Red}_\varepsilon(S)$ is an intersection of sets which by Lemma 4.5 are R-modules that are mapped into themselves by every reduction, it follows that $\mathrm{Red}(S)$ shares these properties.

Next consider t^S. It was shown in Lemma 4.3 that there for every $a \in \mathrm{Per}(S)$ exists some $b \in \mathrm{Irr}(S)$ which is a candidate for being $t^S(a)$, but what about uniqueness? One may observe that if $a \in \mathrm{Red}(S)$ and $b_1, b_2 \in \mathrm{Irr}(S)$ are such that there for every $\varepsilon \in \widehat{\mathcal{O}}$ exist $t_1, t_2 \in T(S)$ such that $t_1(a)$ is stuck in $b_1 + \varepsilon$ and $t_2(a)$ is stuck in $b_2 + \varepsilon$, then by ε-unique reducibility $b_1 - b_2 \in \varepsilon$ for every $\varepsilon \in \widehat{\mathcal{O}}$. Hence $b_1 = b_2$ as $\bigcap_{\varepsilon \in \widehat{\mathcal{O}}} \varepsilon = \{0\}$ and thus t^S is well-defined. The same argument for $(t_1 \circ t)(a)$ and $t_2(a)$ being stuck in $b_1 + \varepsilon$ and $b_2 + \varepsilon$ respectively demonstrates that $(t^S \circ t)(a) = t^S(a)$ for all $t \in T(S)$ and $a \in \mathrm{Red}(S)$. Since an irreducible element b is always stuck in every neighbourhood of itself, it follows that $t^S(b) = b$ for all $b \in \mathrm{Irr}(S)$.

Now let $a_1, a_2 \in \mathrm{Red}(S)$ be given and consider the matter of whether $t^S(a_1 + a_2) = t^S(a_1) + t^S(a_2)$. Let $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary. By definition of t^S there exists some $t_1 \in T(S)$ such that $t_1(a_1)$ is stuck in $t^S(a_1) + \varepsilon$. Since $a_2 \in \mathrm{Per}(S)$ there exists some $t_2 \in T(S)$ and $b_2 \in \mathrm{Irr}(S)$ such that $(t_2 \circ t_1)(a_2)$ is stuck in $b_2 + \varepsilon$, and by unique reducibility of a_2 it follows that $b_2 + \varepsilon =$

$t^S(a_2)+\varepsilon$. Hence $(t_2\circ t_1)(a_1+a_2)$ is stuck in $t^S(a_1)+t^S(a_2)+\varepsilon$, and then by unique reducibility of a_1+a_2 it follows that $t^S(a_1+a_2)+\varepsilon = t^S(a_1)+t^S(a_2)+\varepsilon$. Thus $t^S(a_1+a_2) = t^S(a_1)+t^S(a_2)$ by the arbitrariness of ε.

Next consider the matter of whether a continuous group homomorphism $r\colon \overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ that commutes with all reductions will map $\mathrm{Red}(S)$ into itself and commute with t^S. Let $a \in \mathrm{Red}(S)$ be given and $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary. Let $b \in \mathrm{Irr}(S)$ and $t_1 \in T(S)$ such that $t_1\big(r(a)\big)$ is stuck in $b+\varepsilon$ be arbitrary. By the continuity of r there exists some $\delta \in \widehat{\mathcal{O}}$ such that $r(\delta) \subseteq \varepsilon$; let $t_2 \in T(S)$ be such that $t_2\big(t_1(a)\big)$ is stuck in $t^S(a)+\delta$. Since

$$b - r\big(t^S(a)\big) = b - (t_2\circ t_1\circ r)(a) + r\big((t_2\circ t_1)(a) - t^S(a)\big) \in \varepsilon + r(\delta) = \varepsilon$$

it follows that $r(a)$ is ε-uniquely reducible, and by the arbitrariness of ε that it is uniquely reducible. It is furthermore persistently reducible by Lemma 4.3, and hence an element of $\mathrm{Red}(S)$. Finally $t^S\big(r(a)\big) = r\big(t^S(a)\big)$ since it was in neighbourhoods of $r\big(t^S(a)\big)$ that images of $r(a)$ could get stuck.

Last, it should be verified that $\ker t^S \subseteq \mathcal{I}(S)$. Let $a \in \ker t^S$ be given. Let $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary. There exists some $t \in T(S)$ such that $t(a)$ is stuck in $t^S(a)+\varepsilon = \varepsilon$, and hence $a - t(a) \in a - \varepsilon$ on one hand and $a - t(a) \in \big\{\, b - t(b) \,\big|\, b \in \overline{\mathcal{M}} \,\big\} \subseteq \mathcal{I}(S)$ on the other, i.e., a is a limit point of $\mathcal{I}(S)$. Since $\mathcal{I}(S)$ is topologically closed by definition, $a \in \mathcal{I}(S)$. □

A third property that $\mathrm{Per}_\varepsilon(S)$ and $\mathrm{Red}_\varepsilon(S)$ should possess is to be topologically closed, but this does *not* happen automatically. A minimal example of a situation where $\mathrm{Per}_\varepsilon(S)$ is not topologically closed can be constructed on the formal power series foundation $\mathcal{M} = \mathbb{Z}[\mathsf{a}]$, $\overline{\mathcal{M}} = \mathbb{Z}[\![\mathsf{a}]\!]$, and $\overline{B_n} = \mathsf{a}^n\overline{\mathcal{M}}$, if one picks as $T_1(S) = \{t_n\}_{n=1}^\infty$ where

$$t_n(\mathsf{a}^m) = \begin{cases} \mathsf{a}^{m-1} & \text{if } m = n, \\ \mathsf{a}^m & \text{otherwise} \end{cases}$$

for all $n \geqslant 1$ and $m \geqslant 0$. The problematic trait of this set of reductions is that $(t_1\circ\ldots\circ t_n)(\mathsf{a}^n) = 1 \notin \overline{B_1}$ for any $n \geqslant 1$, although the initial $\mathsf{a}^n \in \overline{B_n}$. This means no proper series $a \in \overline{\mathcal{M}} \setminus \mathcal{M}$ is ever stuck in any set of the form $b+\varepsilon$, and consequently no such element can ever be persistently reducible. It follows that in this case $\mathrm{Red}(S) = \mathrm{Per}(S) = \mathcal{M}$, which is rather small compared to the closure $\overline{\mathcal{M}}$.

On a conceptual level, what breaks down in this example is the principle that series truncation produces a useful approximation. Truncation works for a fixed reduction $t \in T(S)$ — in order to determine $t(a)$ up to a certain number n of terms (i.e., in order to identify $t(a) + \overline{B_n}$) it is sufficient to determine $t(b)$ where b is truncated to some number m of terms; this is the claim that t is continuous — but it may fail when the reduction is not fixed. What one would want is therefore a bound m on the number of terms that

must be taken into account that works for all reductions, and as it happens the property that such a bound exists has a name that is well known in analysis.

Definition 4.8. A set F of group homomorphisms $\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ is said to be **equicontinuous** if there for every $\varepsilon \in \widehat{\mathcal{O}}$ exists some $\delta \in \widehat{\mathcal{O}}$ such that for all $f \in F$ it holds that $f(\delta) \subseteq \varepsilon$.

Returning to the example, one may observe that $T_1(S) = \{t_n\}_{n=1}^{\infty}$ actually is equicontinuous (for $\varepsilon = \overline{B_n}$ take $\delta = \overline{B_{n+1}}$), but what matters is that $T(S)$ is not: no matter how small an a^m may be, there is always a composition of simple reductions that magnifies it to something outside every $\varepsilon \in \widehat{\mathcal{O}}$. On the contrapositive side, when $T(S)$ is equicontinuous then all the sets defined in this section become topologically closed.

Lemma 4.9. *If $T(S)$ is equicontinuous then $t^S\colon \mathrm{Red}(S) \longrightarrow \mathrm{Irr}(S)$ is continuous and furthermore the sets $\mathrm{Per}(S)$, $\mathrm{Per}_\varepsilon(S)$, $\mathrm{Red}(S)$, and $\mathrm{Red}_\varepsilon(S)$ are, for all $\varepsilon \in \widehat{\mathcal{O}}$, topologically closed in $\overline{\mathcal{M}}$.*

Proof. Since t^S by Lemma 4.7 is a group homomorphism, it suffices to show that it is continuous at 0. Let $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary. Let $\delta \in \widehat{\mathcal{O}}$ be such that $t(\delta) \subseteq \varepsilon$ for all $t \in T(S)$; in other words every $a \in \delta$ is stuck in ε. Any $a \in \mathrm{Red}(S) \cap \delta$ thus satisfies $t^S(a) \in \varepsilon$, and hence t^S is continuous at 0.

Now let $\varepsilon \in \widehat{\mathcal{O}}$ be given, and let $\delta \in \widehat{\mathcal{O}}$ be such that $t(\delta) \subseteq \varepsilon$ for all $t \in T(S)$. Let $a \in \overline{\mathcal{M}}$ be an arbitrary limit point of $\mathrm{Per}_\varepsilon(S)$, and let $c \in \mathrm{Per}_\varepsilon(S)$ be such that $c - a \in \delta$. Let $t_1 \in T(S)$ be arbitrary. By persistent reducibility of c there exists some $t_2 \in T(S)$ and $b \in \mathrm{Irr}(S)$ such that $(t_2 \circ t_1)(c)$ is stuck in $b + \varepsilon$, i.e., $(t_3 \circ t_2 \circ t_1)(c) \in b + \varepsilon$ for all $t_3 \in T(S)$. By equicontinuity $(t_3 \circ t_2 \circ t_1)(c) - (t_3 \circ t_2 \circ t_1)(a) \in \varepsilon$ and hence $(t_3 \circ t_2 \circ t_1)(a) \in b + \varepsilon$ as well, which means $(t_2 \circ t_1)(a)$ is stuck in $b + \varepsilon$. By the arbitrariness of t_1 it follows that $a \in \mathrm{Per}_\varepsilon(S)$, and hence that set must be topologically closed. $\mathrm{Per}(S)$ is thus known to be the intersection of a family of topologically closed sets, which implies that it too is closed.

To show that $\mathrm{Red}_\varepsilon(S)$ is topologically closed, let $a \in \mathrm{Per}(S)$ be a limit point of $\mathrm{Red}_\varepsilon(S)$. Let $t_1, t_2 \in T(S)$ and $b_1, b_2 \in \mathrm{Irr}(S)$ be arbitrary such that $t_i(a)$ is stuck in $b_i + \varepsilon$ for $i = 1, 2$. Let $c \in \mathrm{Red}_\varepsilon(S)$ be such that $c - a \in \delta$. Since $t(c) - t(a) \in \varepsilon$ for all $t \in T(S)$, it follows that $t_i(c)$ is also stuck in $b_i + \varepsilon$ for $i = 1, 2$, and hence $b_1 - b_2 \in \varepsilon$ by the ε-unique reducibility of c, whence a is ε-uniquely reducible. All limit points of $\mathrm{Red}_\varepsilon(S)$ are in $\mathrm{Per}(S)$ and thus $\mathrm{Red}_\varepsilon(S)$ is topologically closed. $\mathrm{Red}(S)$ is similarly now known to be the intersection of a family of topologically closed sets, which implies that it is closed as well. □

From an analytical perspective, the effect of equicontinuity of $T(S)$ is rather drastic — $\mathrm{Irr}(S)$ becomes sticky, in the sense that any $a \in \mathrm{Irr}(S) + \delta$ is stuck in the corresponding $\mathrm{Irr}(S) + \varepsilon$ (and even in $a + \varepsilon$) — so in view of the

ruggedness of the proof of Lemma 4.9, one might wonder whether equicontinuity really is The Right Condition for reaching the end that $\mathrm{Red}(S)$ is closed, but the jury is still out on that one. Looking at the proofs certainly suggests that it should be possible to make do with something weaker, but concrete applications rather tend to end up satisfying the stronger condition that $t(\varepsilon) \subseteq \varepsilon$ for all $\varepsilon \in \widehat{\mathcal{O}}$ and $t \in T(S)$. Right now, the best reason for using equicontinuity is probably that it is well established and fully general; many other conditions which at first may seem to give finer control or be easier to verify are only defined with respect to some additional structure, such as a metric.

The next lemma is the first step towards the Diamond Lemma. In natural language, the first of the two equivalent claims is that all elements of $\overline{\mathcal{M}}$ are uniquely reducible, whereas the second claim is that every element has a unique normal form.

Lemma 4.10. *If $T(S)$ is equicontinuous and all elements of $\overline{\mathcal{M}}$ are persistently reducible then the following claims are equivalent:*

- $\mathrm{Red}(S) = \overline{\mathcal{M}}$.
- $\overline{\mathcal{M}} = \mathrm{Irr}(S) \oplus \mathcal{I}(S)$.

Proof. First assume $\mathrm{Red}(S) = \overline{\mathcal{M}}$. By Lemma 4.7, t^S is a projection of $\overline{\mathcal{M}}$ onto $\mathrm{Irr}(S)$, hence $\overline{\mathcal{M}} = \mathrm{Irr}(S) \oplus \ker t^S$. By the same lemma, $\ker t^S \subseteq \mathcal{I}(S)$. Hence $\overline{\mathcal{M}} = \mathrm{Irr}(S) \oplus \mathcal{I}(S)$ will follow if it can be shown that $\mathcal{I}(S) \subseteq \ker t^S$. For any $t \in T(S)$ and $a \in \overline{\mathcal{M}}$ it follows from Lemma 4.7 that $t^S\big(a - t(a)\big) = 0$ and hence $\big\{\, a - t(a) \,\big|\, a \in \overline{\mathcal{M}} \,\big\} \subseteq \ker t^S$ for any $t \in T(S)$. Since $\ker t^S$ is closed under addition,

$$\ker t^S \supseteq \sum_{t \in T(S)} \big\{\, a - t(a) \,\big|\, a \in \overline{\mathcal{M}} \,\big\},$$

and since it by Lemma 4.9 also is topologically closed, the wanted $\mathcal{I}(S) \subseteq \ker t^S$ has been established. This has proved one half of the equivalence.

For the other half, assume $\overline{\mathcal{M}} = \mathrm{Irr}(S) \oplus \mathcal{I}(S)$. Let $a \in \overline{\mathcal{M}}$ and $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary; it will be shown that a is ε-uniquely reducible. Let $t_1, t_2 \in T(S)$ and $b_1, b_2 \in \mathrm{Irr}(S)$ such that $t_i(a)$ is stuck in $b_i + \varepsilon$ for $i = 1, 2$ be arbitrary. By Lemma 4.2 and the persistent reducibility of $t_1(a)$ and $t_2(a)$, there exist $c_1, c_2 \in \mathrm{Irr}(S)$ such that there for every $\delta \in \widehat{\mathcal{O}}$ exist $u_1, u_2 \in T(S)$ such that $u_i\big(t_i(a)\big)$ is stuck in $c_i + \delta$ for $i = 1, 2$. Clearly $u_1\big(t_1(a)\big) - u_2\big(t_2(a)\big) \in \mathcal{I}(S)$, and since such u_1 and u_2 exist for all $\delta \in \widehat{\mathcal{O}}$ it follows that the limit $c_1 - c_2 \in \mathcal{I}(S)$ as well, but since also $c_1 - c_2 \in \mathrm{Irr}(S)$ and $\mathcal{I}(S) \cap \mathrm{Irr}(S) = \{0\}$ it just so happens that $c_1 = c_2$. For either $i = 1, 2$ this common value is a limit point of a sequence of elements that are stuck in the topologically closed set $b_i + \varepsilon$, and hence $c_1 - b_i \in \varepsilon$. Therefore $b_1 - b_2 \in \varepsilon$ and the two neighbourhoods are the same. By the arbitrariness of b_1 and b_2, the element a is ε-uniquely reducible. □

The final lemma in this section explores a slightly different aspect of the machinery: how the sets change if some simple reductions are removed. It sometimes happens when one is preparing a presentation of an argument involving the diamond lemma that some of the reductions turn out to be redundant, but not all ways of verifying this redundancy are as easy as they may seem. The key condition for establishing equivalence of $T_1(S)$ to $T_1(S')$ — i.e., that all things constructed from the set of simple reductions are the same when the set of simple reductions is $T_1(S')$ as when it is $T_1(S)$ — is that their respective sets $\mathrm{Irr}(S)$ and $\mathrm{Irr}(S')$ of irreducible elements are the same; see Theorem 5.6 for a method of characterising the irreducible elements.

Lemma 4.11. *If $T_1(S') \subseteq T_1(S)$ are such that $\mathrm{Irr}(S') = \mathrm{Irr}(S)$, $\mathrm{Per}(S') = \overline{\mathcal{M}}$, $\mathrm{Red}(S) = \overline{\mathcal{M}}$, and $T(S)$ is equicontinuous, then $\mathrm{Red}(S') = \overline{\mathcal{M}}$ as well and $t^{S'} = t^S$.*

Proof. Let $a \in \overline{\mathcal{M}}$ and $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary. What needs to be shown is that a is ε-uniquely reducible under $T(S')$. Therefore let $b_1, b_2 \in \mathrm{Irr}(S')$ and $t_1, t_2 \in T(S')$ be such that $t_1(a)$ is stuck in $b_1 + \varepsilon$ under $T(S')$ and $t_2(a)$ is stuck in $b_2 + \varepsilon$ under $T(S')$. Let $\delta \in \widehat{\mathcal{O}}$ be such that $t(\delta) \subseteq \varepsilon$ for all $t \in T(S)$. By persistent δ-reducibility of a under $T(S')$ there exist $t_1', t_2' \in T(S')$ and $b_1', b_2' \in \mathrm{Irr}(S')$ such that $t_1'\big(t_1(a)\big) \in b_1' + \delta$ and $t_2'\big(t_2(a)\big) \in b_2' + \delta$. By equicontinuity this implies that $t_1'\big(t_1(a)\big)$ and $t_2'\big(t_2(a)\big)$ are stuck under $T(S)$ in $b_1' + \varepsilon$ and $b_2' + \varepsilon$ respectively. By ε-unique reducibility under $T(S)$ of a this implies that $b_1' - b_2' \in \varepsilon$. Furthermore $b_i - b_i' = b_i - t_i'\big(t_i(a)\big) + t_i'\big(t_i(a)\big) - b_i' \in \varepsilon + \delta \subseteq \varepsilon$ for $i = 1, 2$ and thus $b_1 - b_2 \in \varepsilon$ as well. Hence a is indeed ε-uniquely reducible under $T(S')$, and it follows that $\mathrm{Red}(S') = \overline{\mathcal{M}}$.

Now let $a \in \overline{\mathcal{M}}$ be given and consider the matter of whether $t^{S'}(a) = t^S(a)$. Let $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary and let $\delta \in \widehat{\mathcal{O}}$ be such that $t(\delta) \subseteq \varepsilon$ for all $t \in T(S)$. There exists some $t \in T(S') \subseteq T(S)$ such that $t(a) \in t^{S'}(a) + \delta$, hence $t(a)$ is stuck in $t^{S'}(a) + \varepsilon$ under $T(S)$, and consequently $t^S(a) \in t^{S'}(a) + \varepsilon$. It follows from the arbitrariness of ε that $t^S(a) = t^{S'}(a)$. □

5 The core theorem

At the heart of every diamond lemma lies an induction, and the role that $\mathcal{Y}$ will play is as the domain of that induction, to which end it is necessary to order $\mathcal{Y}$. For many novices, the need to systematically order the monomials is by far the most unfamiliar aspect of working with the diamond lemma (and/or Gröbner basis theory), and the problem of constructing a suitable order can be quite baffling. While this is not the place to give advice on how to attack that problem — see instead [7] for tips on this, in particular for issues regarding how the order interacts with the topology — it will still become necessary to reason about orders and their relations to other structures. For that end, it helps to introduce a bit of notation for order relations,

which will facilitate discussions that simultaneously involve several orders. The aim is to allow the order to be an ordinary mathematical letter P (or more generally an expression), rather than a fancy symbol like $\succ$.

The basic claim one can make (with respect to an order relation P) about a pair (μ, ν) of elements is that they are related by this relation. The usual formal interpretation of this is that $(\mu, \nu) \in P$, but notationally it is more convenient to write something like '$\mu \leqslant \nu$ in P', to clarify that this is a non-strict inequality and that μ is on the "small side". Using that one then defines

$$\begin{aligned}
\mu \geqslant \nu \text{ in } P \quad &\Longleftrightarrow \quad \nu \leqslant \mu \text{ in } P, \\
\mu < \nu \text{ in } P \quad &\Longleftrightarrow \quad \mu \leqslant \nu \text{ in } P \text{ and } \nu \not\leqslant \mu \text{ in } P, \\
\mu > \nu \text{ in } P \quad &\Longleftrightarrow \quad \nu < \mu \text{ in } P, \\
\mu \sim \nu \text{ in } P \quad &\Longleftrightarrow \quad \mu \leqslant \nu \text{ in } P \text{ and } \nu \leqslant \mu \text{ in } P.
\end{aligned}$$

If P is a partial order then $\mu \sim \nu$ in P is the same thing as $\mu = \nu$, but if P is a more general quasi-order then this need not be the case. It is often convenient to construct a complex partial order by a sequence of refinements of some simpler quasi-order. Not every partial order will usefully support inductions however, so an additional property is needed.

Definition 5.1. A binary relation P on $\mathcal{Y}$ is said to satisfy the **topological descending chain condition** (or to be **TDCC** for short) if $\lim_{n\to\infty} \mu_n = 0$ for every infinite sequence $\{\mu_n\}_{n=0}^{\infty} \subseteq \mathcal{Y}$ such that $\mu_n > \mu_{n+1}$ in P for all $n \in \mathbb{N}$.

An informal phrasing of the ordinary descending chain condition (DCC) is 'there is no infinite strictly descending chain', although it is important to observe that 'descending' here implicitly requires that the chain elements are indexed. An index-free formulation of the DCC is 'every nonempty subset has a minimal element', and when this definition is given one usually speaks about the order being *well-founded* (which is thus a synonym of DCC). Note that asking for a *minimal* element is weaker than asking for a *minimum* element; the latter would give rise to a well-order, which in particular is always a total order.

The next lemma gives the precise form of an induction over $\mathcal{Y}$; condition (i) provides the induction base, whereas the verification of condition (ii) is the induction step.

Lemma 5.2. *Let P be a partial order on $\mathcal{Y}$ that is TDCC. If $Z \subseteq \mathcal{Y}$ is such that:*

(i) *there exists an $\varepsilon \in \widehat{\mathcal{O}}$ such that $\varepsilon \cap \mathcal{Y} \subseteq Z$, and*

(ii) *if $\mu \in \mathcal{Y}$ is such that $\{\, \nu \in \mathcal{Y} \,|\, \nu < \mu \text{ in } P \,\} \subseteq Z$ then $\mu \in Z$;*

then $Z = \mathcal{Y}$.

Proof. Let Z be an arbitrary proper subset of $\mathcal{Y}$ which satisfies (ii); it will be shown that Z does not satisfy (i). To see this, let $\mu_0 \in \mathcal{Y} \setminus Z$. For any $\mu_n \in \mathcal{Y}\setminus Z$ there must exist a $\mu_{n+1} \in \mathcal{Y}\setminus Z$ such that $\mu_{n+1} < \mu_n$ in P, because if that was not the case then all $\nu \in \mathcal{Y}$ which satisfy $\nu < \mu_n$ in P would also satisfy $\nu \in Z$, and hence by (ii) $\mu_n \in Z$, which would be a contradiction. Thus there exists an infinite P-descending sequence $\{\mu_n\}_{n=1}^{\infty} \subseteq \mathcal{Y} \setminus Z$, and hence by TDCC $\lim_{n\to\infty} \mu_n = 0$. In other words there exists for each $\varepsilon \in \widehat{\mathcal{O}}$ an integer N such that $\mu_n \in \varepsilon$ for all $n \geqslant N$, and thus $\varepsilon \cap \mathcal{Y} \ni \mu_N \notin Z$. Hence Z does not satisfy (i). □

For such inductions to be useful in the present context, it is however necessary that the reductions used comply with the order.

Definition 5.3. If P is a binary relation on $\mathcal{Y}$ and $\mu \in \mathcal{Y}$ then $\mathrm{DSM}(\mu, P)$ denotes the least topologically closed R-module of $\overline{\mathcal{M}}$ which contains all $\nu \in \mathcal{Y}$ such that $\nu < \mu$ in P, i.e.,

$$\mathrm{DSM}(\mu, P) = \mathrm{Cspan}\big(\{\, \nu \in \mathcal{Y} \,|\, \nu < \mu \text{ in } P \,\}\big). \tag{5.1}$$

The set $\mathrm{DSM}(\mu, P)$ is called the **down-set module** of μ with respect to P.

A reduction $t \in T(S)$ is said to be **compatible** with the relation P if $t(\mu) \in \{\mu\} \cup \mathrm{DSM}(\mu, P)$ for all $\mu \in \mathcal{Y}$. A set of reductions is said to be compatible with P if all its elements are compatible with P.

An element $g \in \overline{\mathcal{M}}$ is said to be P**-monic** if there exists some $\mu \in \mathcal{Y}$ such that $g - \mu \in \mathrm{DSM}(\mu, P)$. A subset of $\overline{\mathcal{M}}$ is said to be P-monic if all elements in it are P-monic.

Down-set is a term from poset theory, but the standard name there for this concept is *ideal* rather than down-set. That terminology has however been avoided so that no confusion with the ring-theoretic ideal concept will arise.

That an arbitrary map $\overline{\mathcal{M}} \longrightarrow \overline{\mathcal{M}}$ should be compatible with some order relation P is a rather strong condition, but for reductions it is often something which comes naturally. For a simple reduction $t_{\mu\mapsto a}$ satisfying (3.1), it boils down to the condition that $a \in \mathrm{DSM}(\mu, P)$, and when that is the case then $\mu - a$ is P-monic. Conversely, if (3.1) for any $(\mu, a) \in \mathcal{Y} \times \overline{\mathcal{M}}$ defines a continuous homomorphism $t_{\mu\mapsto a}$ that commutes with elements of R, then every P-monic $g \in \overline{\mathcal{M}}$ gives rise to some map $t_{\mu\mapsto\mu-g}$ that is compatible with P. Gröbner basis theory preaches an extreme form of this, where the normal state of things is that the leading monomial μ is split off from a basis element g every time something is to be reduced modulo g (although it is recognised that caching μ with g can improve performance). This approach is facilitated by the fact that Gröbner basis theory normally

only considers total orders, as that guarantees that there always is a unique leading monomial to split off.

In more general cases, compatibility is often something one arrives at indirectly. When preparing to apply the diamond lemma, one often starts with some reductions $t_{\mu\mapsto a}$ that one wants to use, and faces the task of constructing some P with which these would be compatible. (This P will also have to satisfy some other conditions, in particular the TDCC and in most cases some variant of (6.4), which constrains the possibilities quite a lot.) If the given reductions do not generate all the wanted congruences, then the next step is to find a set of P-monic generators which cover the rest, and then make additional simple reductions from these. At each step one's choices are restricted by the need to ensure compatibility further on, but when the set-up is complete it is usually a trivial matter to verify the compatibility of simple reductions. The next lemma then extends this result to general reductions.

Lemma 5.4. *If P is a partial order on $\mathcal{Y}$ with which $t \in T(S)$ is compatible, then for all $\mu \in \mathcal{Y}$ and $b \in \mathrm{DSM}(\mu, P)$ it holds that $t(b) \in \mathrm{DSM}(\mu, P)$. If P is a partial order on $\mathcal{Y}$ with which $t_1, t_2 \in T(S)$ are compatible, then $t_2 \circ t_1$ is compatible with P as well. If $T_1(S)$ is compatible with a partial order P on $\mathcal{Y}$ then the whole of $T(S)$ is compatible with P.*

Proof. For the first claim, consider first some special b, and then generalise following the characterisation (5.1) of $\mathrm{DSM}(\mu, P)$. If $b = \nu \in \mathcal{Y}$ satisfies $\nu < \mu$ in P then $\mathrm{DSM}(\nu, P) \subseteq \mathrm{DSM}(\mu, P)$ and consequently $t(\nu) \in \{\nu\} \cup \mathrm{DSM}(\nu, P) \subseteq \mathrm{DSM}(\mu, P)$ as claimed. If $r \in R^*$ is any finite composition of elements of R then $t\big(r(\nu)\big) = r\big(t(\nu)\big) \in r\big(\mathrm{DSM}(\mu, P)\big) \subseteq \mathrm{DSM}(\mu, P)$, thus extending the result to b on the form $r(\nu)$. If $b_1, b_2 \in \mathrm{DSM}(\mu, P)$ are such that $t(b_1), t(b_2) \in \mathrm{DSM}(\mu, P)$ then clearly $t(b_1 - b_2) \in \mathrm{DSM}(\mu, P)$ as well, and since this establishes that the set of b for which the result holds is a group it follows that the result holds for arbitrary $b \in \mathrm{Span}\big(\{\, \nu \in \mathcal{Y} \,|\, \nu < \mu \text{ in } P \,\}\big)$. Finally if $\{b_n\}_{n=1}^{\infty} \subseteq \mathrm{DSM}(\mu, P)$ are such that $t(b_n) \in \mathrm{DSM}(\mu, P)$ then $t(b) \in \mathrm{DSM}(\mu, P)$ also for $b = \lim_{n\to\infty} b_n$ by continuity, and thus the result holds for all $b \in \mathrm{DSM}(\mu, P)$.

For the second claim, let $\mu \in \mathcal{Y}$ be arbitrary. If $t_1(\mu) = \mu$ then $(t_2 \circ t_1)(\mu) = t_2(\mu) \in \{\mu\} \cup \mathrm{DSM}(\mu, P)$ by the compatibility of t_2 with P. Otherwise $t_1(\mu) \in \mathrm{DSM}(\mu, P)$ and thus $(t_2 \circ t_1)(\mu) \in \mathrm{DSM}(\mu, P)$ by the first claim. Hence $t_2 \circ t_1$ is compatible with P. The third claim immediately follows from the second and the observation that the identity map id is compatible with all relations. □

An intuitive picture which might be useful is to think of the down-set module of μ as a sort of cone with μ at the apex. This picture is deceiving insofar as it represents entire R-modules of the form $\mathrm{Span}\big(\{\mu\}\big)$ as single points and does not even begin to consider the topological structure, but it is

nonetheless very much to the point. In that picture, one might interpret the above lemma as saying compatible reductions cannot map elements inside a cone to elements outside it. This is similar to how the sets $\mathrm{Per}_\varepsilon(S)$ of persistently ε-reducible elements behave with respect to reductions, and indeed the next lemma makes use of down-set modules in showing that $\mathrm{Per}_\varepsilon(S)$ is the whole of $\overline{\mathcal{M}}$.

Lemma 5.5. *Assume $T(S)$ is equicontinuous and compatible with some partial order P on $\mathcal{Y}$. If P satisfies the topological descending chain condition then $\mathrm{Per}(S) = \overline{\mathcal{M}}$.*

Proof. Let $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary. Let Z be the set of all elements of $\mathcal{Y}$ which are persistently ε-reducible. By equicontinuity there exists some $\delta \in \widehat{\mathcal{O}}$ such that $t(\delta) \subseteq \varepsilon$ for all $t \in T(S)$. Hence all $\mu \in \mathcal{Y} \cap \delta$ belong to Z, since these satisfy $t(\mu) \in \varepsilon$ for all $t \in T(S)$. These elements constitute the base for the induction, fulfilling condition (i) of Lemma 5.2.

For the induction step, consider some arbitrary $\mu \in \mathcal{Y}$. Assume that all $\nu \in \mathcal{Y}$ such that $\nu < \mu$ in P satisfy $\nu \in Z$; it will now be shown that this implies $\mu \in Z$. To that end, let $t_1 \in T(S)$ be given and try to find some $t_2 \in T(S)$ and $a \in \mathrm{Irr}(S)$ such that $t_2\big(t_1(\mu)\big)$ is stuck in $a + \varepsilon$. It is useful to observe that the induction hypothesis, by Lemmas 4.3 and 4.9, implies $\mathrm{DSM}(\mu, P) \subseteq \mathrm{Per}_\varepsilon(S)$.

Depending on μ and t_1, there are three cases that can occur. If $\mu \in \mathrm{Irr}(S)$ then $t_1(\mu) = \mu$ is stuck in $\mu + \varepsilon$ and hence $\mu \in Z$. If $t_1(\mu) \neq \mu$ (and hence $\mu \notin \mathrm{Irr}(S)$) then by compatibility of t_1 with P it follows that $t_1(\mu) \in \mathrm{DSM}(\mu, P) \subseteq \mathrm{Per}_\varepsilon(S)$ and consequently by this persistent ε-reducibility of $t_1(\mu)$ there exist $a \in \mathrm{Irr}(S)$ and $t_2 \in T(S)$ such that $t_2\big(t_1(\mu)\big)$ is stuck in $a + \varepsilon$. Finally, if $\mu \notin \mathrm{Irr}(S)$ but $t_1(\mu) = \mu$ then there still exists some $t_2' \in T(S)$ such that $t_2'(\mu) \neq \mu$ and thus $t_2'(\mu) \in \mathrm{DSM}(\mu, P) \subseteq \mathrm{Per}_\varepsilon(S)$. As before there now exist $t_2'' \in T(S)$ and $a \in \mathrm{Irr}(S)$ such that $t_2''\big(t_2'(\mu)\big)$ is stuck in $a + \varepsilon$, whence for $t_2 = t_2'' \circ t_2'$ one finds that $t_2\big(t_1(\mu)\big)$ is stuck in $a + \varepsilon$. Either way, $\mu \in Z$ by the arbitrariness of t_1, which completes the induction step.

All conditions for Lemma 5.2 are now fulfilled and hence $Z = \mathcal{Y}$. By Lemmas 4.3 and Lemma 4.9, $\mathrm{Per}_\varepsilon(S) = \overline{\mathcal{M}}$. By the arbitrariness of ε, it then follows that $\mathrm{Per}(S) = \bigcap_{\varepsilon \in \widehat{\mathcal{O}}} \mathrm{Per}_\varepsilon(S) = \overline{\mathcal{M}}$ as well. □

The same conditions also suffice for giving an explicit description of $\mathrm{Irr}(S)$. It is not unusual that one can quickly establish this result also through more elementary arguments, but for complicated set-ups it is convenient to have a proof relying on (a subset of) the conditions of Theorem 5.11.

Theorem 5.6. *Assume $T(S)$ is equicontinuous and compatible with some partial order P on $\mathcal{Y}$. If P satisfies the topological descending chain condition then*

$$\mathrm{Irr}(S) = \mathrm{Cspan}\Big(\big\{\, \mu \in \mathcal{Y} \,\big|\, t(\mu) = \mu \text{ for all } t \in T_1(S) \,\big\}\Big). \tag{5.2}$$

Proof. Let W be the set of irreducible elements of $\mathcal{Y}$. It follows from Lemma 3.5 that the left hand side $\mathrm{Irr}(S)$ of (5.2) contains the right hand side $\mathrm{Cspan}(W)$. The reverse inclusion will be established by demonstrating that $\mathrm{Irr}(S) \subseteq \mathrm{Span}(W) + \varepsilon$ for all $\varepsilon \in \widehat{\mathcal{O}}$.

Let $\varepsilon \in \widehat{\mathcal{O}}$ be given. Let $\delta \in \widehat{\mathcal{O}}$ be such that $t(\delta) \subseteq \varepsilon$ for all $t \in T(S)$. Let N be the set of those $a \in \overline{\mathcal{M}}$ such that there for every $t_1 \in T(S)$ exists some $t_2 \in T(S)$ such that for every $t_3 \in T(S)$ it holds that $(t_3 \circ t_2 \circ t_1)(a) \in \mathrm{Span}(W) + \varepsilon$. It will now be shown that $N = \overline{\mathcal{M}}$.

To see that N is a subgroup of $\overline{\mathcal{M}}$, let $a_1, a_2 \in N$ be given. Let $t_1 \in T(S)$ be arbitrary. There exists some $t_2 \in T(S)$ such that $(t_3 \circ t_2 \circ t_1)(a_1) \in \mathrm{Span}(W) + \varepsilon$ for every $t_3 \in T(S)$. There also exists some $t_4 \in T(S)$ such that $\big(t_3 \circ t_4 \circ (t_2 \circ t_1)\big)(a_2) \in \mathrm{Span}(W) + \varepsilon$ for every $t_3 \in T(S)$. Hence $\big(t_3 \circ (t_4 \circ t_2) \circ t_1\big)(a_1 - a_2) = \big((t_3 \circ t_4) \circ t_2 \circ t_1\big)(a_1) - \big(t_3 \circ t_4 \circ (t_2 \circ t_1)\big)(a_2) \in \mathrm{Span}(W) + \varepsilon$ for every $t_3 \in T(S)$, and thus $a_1 - a_2 \in N$ by the arbitrariness of t_1. Elements of R map N into itself because they commute with all reductions and map $\mathrm{Span}(W) + \varepsilon$ into itself, hence N is an R-module. N is topologically closed because any $t \in T(S)$ maps $N + \delta$ into $t(N) + \varepsilon$; if $a \in \overline{N}$ then there is some $a' \in N \cap (a + \delta)$, hence for any $t_1 \in T(S)$ there exists some $t_2 \in T(S)$ such that $(t_3 \circ t_2 \circ t_1)(a') \in \mathrm{Span}(W) + \varepsilon$ for all $t_3 \in T(S)$, and thus $(t_3 \circ t_2 \circ t_1)(a) \in (t_3 \circ t_2 \circ t_1)(a') + \varepsilon \subseteq \mathrm{Span}(W) + \varepsilon + \varepsilon$.

The proof that $\mathcal{Y} \subseteq N$ is done by induction. As usual, $\mathcal{Y} \cap \delta \subseteq \delta \subseteq N$ since $t(\delta) \subseteq 0 + \varepsilon$ for all $t \in T(S)$. For the induction step, let $\mu \in \mathcal{Y}$ be arbitrary and assume $\nu \in N$ for all $\nu \in \mathcal{Y}$ such that $\nu < \mu$ in P. If $\mu \in W$ then $t(\mu) \in W$ for all $t \in T(S)$ and hence $\mu \in N$. Otherwise let $t_1 \in T(S)$ be given. If $t_1(\mu) = \mu$ then let $t_2' \in T_1(S)$ be such that $t_2'(\mu) \neq \mu$, otherwise let $t_2' = \mathrm{id}$. Since $(t_1' \circ t_1)(\mu) \neq \mu$ it follows from compatibility, the induction hypothesis, and the previous paragraph that $(t_2' \circ t_1)(\mu) \in \mathrm{DSM}(\mu, P) \subseteq N$. Hence there exists some $t_2 \in T(S)$ such that any $t_3 \in T(S)$ satisfies $(t_3 \circ t_2 \circ t_2' \circ t_1)(\mu) \in \mathrm{Span}(W) + \varepsilon$ and thus $\mu \in N$. Since P is TDCC, the conclusions $\mathcal{Y} \subseteq N$ and $N = \overline{\mathcal{M}}$ follow.

Finally, let $a \in \mathrm{Irr}(S)$ be arbitrary. Since $a \in N$ there exists some $t \in T(S)$ such that $t(a) \in \mathrm{Span}(W) + \varepsilon$, but $t(a) = a$ by irreducibility. Hence $\mathrm{Irr}(S) \subseteq \mathrm{Span}(W) + \varepsilon$ by the arbitrariness of a, and it follows that $\mathrm{Irr}(S) \subseteq \bigcap_{\varepsilon \in \widehat{\mathcal{O}}} \big(\mathrm{Span}(W) + \varepsilon\big) = \mathrm{Cspan}(W)$. □

The $\mathrm{Red}(S) = \overline{\mathcal{M}}$ counterpart of Lemma 5.5 is Lemma 5.8, but unique reducibility requires another condition, wherein the following definition is handy.

Definition 5.7. Let P be a binary relation on $\mathcal{Y}$. Then for any $\mu \in \mathcal{Y}$, define

$$\mathrm{DIS}(\mu, P, S) = \mathrm{Cspan}\Big(\big\{\, \nu - t(\nu) \;\big|\; \nu < \mu \text{ in } P, t \in T_1(S) \,\big\}\Big). \tag{5.3}$$

Also write $a \equiv b \pmod{S < \mu \text{ in } P}$ (read "a is congruent to b mod S less μ in P") as a shorthand for $a - b \in \mathrm{DIS}(\mu, P, S)$.

The etymology of this DIS notation is "Down-set $\mathcal{I}(S)$ Section", even though one cannot in general interpret $\mathrm{DIS}(\mu, P, S)$ as being synonymous to $\mathrm{DSM}(\mu, P) \cap \mathcal{I}(S)$. It is clear that $\mathrm{DIS}(\mu, P, S) \subseteq \mathcal{I}(S)$, and if $T(S)$ is compatible with P then $\mathrm{DIS}(\mu, P, S) \subseteq \mathrm{DSM}(\mu, P)$ by Lemma 5.4, but for equality with the intersection to hold one pretty much have to fulfil the conditions of Theorem 5.11. It is however not so interesting to exactly map the extent of this module; one rather seeks to prove that particular elements of $\overline{\mathcal{M}}$ belong to it by exhibiting explicit expressions for them. Calculations are often convenient to express in the form $a \equiv a_1 \equiv \cdots \equiv a_n \equiv 0 \pmod{S < \mu \text{ in } P}$.

Lemma 5.8. *Let P be a partial order on $\mathcal{Y}$. Assume $T_1(S)$ is such that for all $\mu \in \mathcal{Y}$ and simple reductions $t_1, t_2 \in T_1(S)$ that act nontrivially on μ it holds that*

$$t_1(\mu) - t_2(\mu) \in \mathrm{DIS}(\mu, P, S). \tag{5.4}$$

If furthermore P satisfies the topological descending chain condition and $T(S)$ is equicontinuous and compatible with P, then $\mathrm{Red}(S) = \overline{\mathcal{M}}$.

Proof. It will be shown by induction over $\mathcal{Y}$ that $\mathrm{Red}_\varepsilon(S) = \overline{\mathcal{M}}$ for all $\varepsilon \in \widehat{\mathcal{O}}$. Observe that $\mathrm{Per}(S) = \overline{\mathcal{M}}$ by Lemma 5.5; hence it is sufficient to prove that all elements of $\overline{\mathcal{M}}$ are ε-uniquely reducible. Let $\varepsilon \in \widehat{\mathcal{O}}$ be given and let $\delta \in \widehat{\mathcal{O}}$ be such that $t(\delta) \subseteq \varepsilon$ for all $t \in T(S)$. The induction hypothesis is that the $\mu \in \mathcal{Y}$ under consideration satisfies $\mu \in \mathrm{Red}_\varepsilon(S)$. The induction hypothesis clearly holds for all $\mu \in \mathcal{Y}$ such that $\mu \in \delta$, since $t(\mu) \in \varepsilon$ for all such μ and all $t \in T(S)$. This has laid the base for the induction.

For the induction step, assume that $\nu \in \mathrm{Red}_\varepsilon(S)$ for all $\nu \in \mathcal{Y}$ such that $\nu < \mu$ in P, and consider μ. By Lemmas 4.5 and 4.9, this assumption is equivalent to $\mathrm{DSM}(\mu, P) \subseteq \mathrm{Red}_\varepsilon(S)$. Let $t_1, t_2 \in T(S)$ and $b_1, b_2 \in \mathrm{Irr}(S)$ be arbitrary elements such that $t_1(\mu)$ is stuck in $b_1 + \varepsilon$ and $t_2(\mu)$ is stuck in $b_2 + \varepsilon$. The problem now is to show that $b_1 + \varepsilon = b_2 + \varepsilon$. A trivial case occurs if t_1 or t_2 acts trivially on μ; it can be assumed without loss of generality that t_2 acts trivially. In this case $\mu = t_2(\mu)$ is already known to be stuck in $b_2 + \varepsilon$, and hence $t_1(\mu)$ is stuck there as well.

With that taken care of, it can be assumed that t_1 and t_2 both act nontrivially on μ. Thus there exist $t_{1a}, t_{2a} \in T_1(S)$ and $t_{1b}, t_{2b} \in T(S)$ such that: $t_1(\mu) = t_{1b}\big(t_{1a}(\mu)\big)$, $t_2(\mu) = t_{2b}\big(t_{2a}(\mu)\big)$, $t_{1a}(\mu) \neq \mu$ and $t_{2a}(\mu) \neq \mu$. By persistent reducibility there also exists some $t_3 \in T(S)$ and $b_3 \in \mathrm{Irr}(S)$ such that $t_3\big(t_{1a}(\mu) - t_{2a}(\mu)\big)$ is stuck in $b_3 + \varepsilon$. By ε-unique reducibility of $t_{1a}(\mu)$ there exists some $t_{1c} \in T(S)$ such that $(t_{1c} \circ t_3 \circ t_{1a})(\mu)$ is stuck in $b_1 + \varepsilon$, and similarly there exists some $t_{2c} \in T(S)$ such that $(t_{2c} \circ t_{1c} \circ t_3 \circ t_{2a})(\mu)$ is stuck in $b_2 + \varepsilon$. Let $t_4 = t_{2c} \circ t_{1c}$. Since

$$\begin{aligned} b_1 - b_2 = \big(b_1 - (t_4 \circ t_3 \circ t_{1a})(\mu)\big) + \\ + (t_4 \circ t_3)\big(t_{1a}(\mu) - t_{2a}(\mu)\big) + \big((t_4 \circ t_3 \circ t_{2a})(\mu) - b_2\big) \in \\ \in \varepsilon + (b_3 + \varepsilon) + \varepsilon, \end{aligned}$$

it would follow that $b_1 + \varepsilon = b_2 + \varepsilon$ if $b_3 \in \varepsilon$.

By assumption $t_{1a}(\mu) - t_{2a}(\mu) \in \mathrm{DIS}(\mu, P, S)$. Thus there exist $\{\rho_i\}_{i=1}^n \subseteq \mathcal{Y}$, reductions $\{u_i\}_{i=1}^n \subseteq T_1(S)$, and $a_i \in \mathrm{Span}\big(\{\rho_i\}\big)$ for each $i = 1, \dots, n$, such that $\rho_i < \mu$ in P for $i = 1, \dots, n$ and

$$\big(t_{1a}(\mu) - t_{2a}(\mu)\big) - \sum_{i=1}^{n} \big(a_i - u_i(a_i)\big) \in \delta.$$

Now the idea is to construct and consider a reduction that takes each term of this expression to an ε-neighbourhood of its normal form. Let $w_0 = t_3$ and for each $i = 1, \dots, n$: let $v_i, v_i' \in T(S)$ and $c_i, c_i' \in \mathrm{Irr}(S)$ be such that $v_i\big(w_{i-1}(a_i)\big)$ is stuck in some $c_i + \varepsilon$ and $v_i'\big((v_i \circ w_{i-1} \circ u_i)(a_i)\big)$ is stuck in $c_i' + \varepsilon$, then define $w_i = v_i' \circ v_i \circ w_{i-1}$. By ε-unique reducibility of a_i it follows that $c_i + \varepsilon = c_i' + \varepsilon$ and thus $w_n\big(a_i - u_i(a_i)\big) \in \varepsilon$. Hence

$$w_n\big(t_{1a}(\mu) - t_{2a}(\mu)\big) \in w_n\bigg(\sum_{i=1}^{n} \big(a_i - u_i(a_i)\big) + \delta\bigg) = \\ = \sum_{i=1}^{n} w_n\big(a_i - u_i(a_i)\big) + w_n(\delta) \subseteq \varepsilon.$$

Since also $w_n\big(t_{1a}(\mu) - t_{2a}(\mu)\big) \in b_3 + \varepsilon$, it follows that $b_3 \in \varepsilon$ and $b_1 + \varepsilon = b_2 + \varepsilon$. This completes the induction step.

By Lemma 5.2, the induction hypothesis holds for all $\mu \in \mathcal{Y}$, and hence $\mathcal{Y} \subseteq \mathrm{Red}_\varepsilon(S)$, which implies $\mathrm{Red}_\varepsilon(S) = \overline{\mathcal{M}}$. Since ε was arbitrary, $\mathrm{Red}(S) = \bigcap_{\varepsilon > 0} \mathrm{Red}_\varepsilon(S) = \overline{\mathcal{M}}$ as well. □

The next definition helps simplify the main condition (5.4) of Lemma 5.8 to "assume all ambiguities of $T_1(S)$ are resolvable relative to P", which is one of the main equivalent conditions in the diamond lemma.

Definition 5.9. An **ambiguity** of $T_1(S)$ is a triplet (t_1, μ, t_2), where $t_1, t_2 \in T_1(S)$ act nontrivially on $\mu \in \mathcal{Y}$; the ambiguities (t_1, μ, t_2) and (t_2, μ, t_1) are considered equivalent. An ambiguity (t_1, μ, t_2) is said to be **resolvable** if there, for every $\varepsilon \in \widehat{\mathcal{O}}$, exists reductions $t_3, t_4 \in T(S)$ such that $t_3\big(t_1(\mu)\big) - t_4\big(t_2(\mu)\big) \in \varepsilon$. An ambiguity (t_1, μ, t_2) is said to be **resolvable relative to** a binary relation P on $\mathcal{Y}$ if $t_1(\mu) - t_2(\mu) \in \mathrm{DIS}(\mu, P, S)$.

The essential content of the ambiguity concept has been given a bewildering variety of names, where 'ambiguity' is that used by Bergman [2]. The most common term is rather *critical pair*, but there appears to be no consensus on what the elements of the critical pair are. Baader–Nipkow [1, Def. 6.2.1] effectively defines a critical pair to be some $\big(t_1(\mu), t_2(\mu)\big)$ and informally speaks of the ambiguity (t_1, μ, t_2) from which it came as a *fork*. This critical pair terminology would make sense within the present framework, but it cannot completely replace ambiguities, as there is not enough

information in the critical pair to define relative resolvability. In contrast, the definition of 'critical pair' in Gröbner basis theory (a pair of Gröbner basis elements) is technically quite different and cannot be stated in the generic framework, although the essential content is still the same.

The second most common name is probably 'overlap', but although overlap ambiguities are by far the most important ones, there are also important ambiguities which aren't overlaps; the taxonomy of ambiguities is a subject of Section 6. Rarer names still are 'composition' (Shirshov [14] and Bokut [3], hence the alternative name *composition lemma* for the diamond lemma) and 'superposition' (Knuth–Bendix [9]), both of which refer primarily to the μ part of an ambiguity (t_1, μ, t_2).

It should also be pointed out that many of the above concepts presume a certain minimality — the *critical* of 'critical pair' refers to that these are the ones that really need to be checked — whereas the above ambiguity concept has no such restriction. This is because the mechanisms traditionally used to discard some ambiguities as redundant rely on structures not apparent in the basic framework $\big(\mathcal{M}, R, \mathcal{Y}, \mathcal{O}, T_1(S)\big)$, and therefore not available in this generality. Corresponding results for the present setting can be found in Section 6.

It is common to say that a rewriting system S is *confluent* if everything has a unique normal form, but since this by Theorem 5.11 is equivalent to a number of quite different conditions, one shouldn't be surprised if different authors define it differently. Taking [1] as authority, where confluence is defined for reduction relations, one may call $T_1(S)$ *locally confluent* if all ambiguities of $T_1(S)$ are resolvable; this adjusts the traditional definition to allow for topology and take advantage of the R-module structure, but is otherwise a strict interpretation. Global confluence is, assuming persistent reducibility, more directly equivalent to unique reducibility: $T_1(S)$ is globally confluent if there for every $\mu \in \mathcal{Y}$, all $t_1, t_2 \in T(S)$ (*not* only simple reductions), and every $\varepsilon \in \widehat{\mathcal{O}}$ exist $t_3, t_4 \in T(S)$ such that $t_3\big(t_1(\mu)\big) - t_4\big(t_2(\mu)\big) \in \varepsilon$. The two may seem similar, but a proof that local confluence implies global confluence (which essentially is what the original diamond lemma of Newman [12] was all about) requires something like an induction over $\mathcal{Y}$ to go through.

As will become clear in the next section, relative resolvability is more important in the theoretical machinery than plain resolvability, since it more easily lends itself to reasoning about elements of $\overline{\mathcal{M}} \setminus \mathcal{Y}$. On the other hand, plain resolvability is usually a more natural goal to aim for in practical calculations. The next lemma says that it is a sufficient condition also for relative resolvability.

Lemma 5.10. *If $T(S)$ is compatible with the partial order P on $\mathcal{Y}$, then*

$$\mathrm{DIS}(\mu, P, S) = \overline{\sum_{t \in T_1(S)} \left\{ a - t(a) \,\middle|\, a \in \mathrm{DSM}(\mu, P) \right\}} \tag{5.5}$$

$$= \overline{\sum_{t \in T(S)} \left\{ a - t(a) \,\middle|\, a \in \mathrm{DSM}(\mu, P) \right\}} \tag{5.6}$$

for all $\mu \in \mathcal{Y}$ and each resolvable ambiguity of $T_1(S)$ is also resolvable relative to P.

Proof. Let $\mu \in \mathcal{Y}$ be given. It is clear that (5.5) and (5.6) hold with $\subseteq$ inclusions, so what needs to be shown are the $\supseteq$ inclusions. In (5.6), one may observe that any $t \in T(S) \setminus \{\mathrm{id}\}$ decomposes as $t = t_n \circ \cdots \circ t_1$ for $t_1, \ldots, t_n \in T_1(S)$ and that $a - t(a) = \sum_{i=1}^n \big(a_i - t_i(a_i)\big)$ for $a_1 = a$ and $a_{i+1} = (t_i \circ \cdots \circ t_1)(a)$ for $i = 1, \ldots, n-1$. By Lemma 5.4, $a_i \in \mathrm{DSM}(\mu, P)$ for $i = 1, \ldots, n$, and hence

$$\left\{ a - t(a) \,\middle|\, a \in \mathrm{DSM}(\mu, P), t \in T(S) \right\} \subseteq \sum_{t \in T_1(S)} \left\{ a - t(a) \,\middle|\, a \in \mathrm{DSM}(\mu, P) \right\}.$$

In (5.5), one must instead decompose the elements of $\mathrm{DSM}(\mu, P)$. Let $t \in T_1(S)$ and $a \in \mathrm{DSM}(\mu, P)$ be arbitrary. Let $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary. Let $\delta \in \widehat{\mathcal{O}}$ be such that $\delta \subseteq \varepsilon$ and $t(\delta) \subseteq \varepsilon$. By Lemma 2.8, there exist $\nu_1, \ldots, \nu_n \in \mathcal{Y}$ and $r_1, \ldots, r_n \in \pm R^*$ such that $\nu_i < \mu$ in P for $i = 1, \ldots, n$ and $b := \sum_{i=1}^n r_i(\nu_i) \in a + \delta$. Clearly

$$b - t(b) = \sum_{i=1}^n r_i(\nu_i) - t\Big(\sum_{i=1}^n r_i(\nu_i)\Big) = \sum_{i=1}^n r_i\big(\nu_i - t(\nu_i)\big) \in \mathrm{DIS}(\mu, P, S)$$

and since $\big(b - t(b)\big) - \big(a - t(a)\big) \in \varepsilon$ it follows, by the arbitrariness of ε, that $a - t(a)$ is a limit point of $\mathrm{DIS}(\mu, P, S)$. Since this set is topologically closed by definition, $a - t(a) \in \mathrm{DIS}(\mu, P, S)$ and hence

$$\sum_{t \in T_1(S)} \left\{ a - t(a) \,\middle|\, a \in \mathrm{DSM}(\mu, P) \right\} \subseteq \mathrm{DIS}(\mu, P, S)$$

by the arbitrariness of a and t. (5.5) follows.

Let (t_1, μ, t_2) be a resolvable ambiguity of $T_1(S)$, let $a_1 = t_1(\mu)$, and let $a_2 = t_2(\mu)$. Let $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary. Let $t_3, t_4 \in T(S)$ be such that $t_3(a_1) - t_4(a_2) \in \varepsilon$. By Lemma 5.4, $a_1, a_2 \in \mathrm{DSM}(\mu, P)$. Hence $b = \big(a_1 - t_3(a_1)\big) + \big(-a_2 - t_4(-a_2)\big) \in \mathrm{DIS}(\mu, P, S)$ and $(a_1 - a_2) - b = t_3(a_1) - t_4(a_2) \in \varepsilon$. In other words, $a_1 - a_2$ is a limit point of $\mathrm{DIS}(\mu, P, S)$. As above, it follows that $a_1 - a_2 \in \mathrm{DIS}(\mu, P, S)$. □

With that final implication, the big equivalence in the generic diamond lemma is now apparent:

Theorem 5.11. *If $T(S)$ is equicontinuous and compatible with a partial order P on $\mathcal{Y}$ that furthermore satisfies the topological descending chain condition, then the following claims are equivalent:*

(a) *Every ambiguity of $T_1(S)$ is resolvable.*

(a′) *Every ambiguity of $T_1(S)$ is resolvable relative to P.*

(b) *Every element of $\overline{\mathcal{M}}$ is persistently and uniquely reducible, i.e., $\overline{\mathcal{M}} = \mathrm{Red}(S)$.*

(c) *Every element of $\overline{\mathcal{M}}$ has a unique normal form, i.e., $\overline{\mathcal{M}} = \mathrm{Irr}(S) \oplus \mathcal{I}(S)$.*

Proof. By Lemma 5.5, $\mathrm{Per}(S) = \overline{\mathcal{M}}$. Hence (b) and (c) are equivalent by Lemma 4.10. (a) implies (a′) by Lemma 5.10 and (a′) implies (b) by Lemma 5.8. Hence the only thing left to prove is that (b) implies (a).

Let an ambiguity (t_1, μ, t_2) be given. Let $\varepsilon \in \widehat{\mathcal{O}}$ be arbitrary. Since $\mu \in \mathrm{Red}(S)$ there exist $t_3, t_4 \in T(S)$ such that $t_3\big(t_1(\mu)\big)$ and $t_4\big(t_2(\mu)\big)$ are stuck in $t^S(\mu) + \varepsilon$. Hence $t_3\big(t_1(\mu)\big) - t_4\big(t_2(\mu)\big) \in \varepsilon$, and thus the ambiguity is resolvable. □

6 Ambiguities

Most applications of a diamond lemma has as one of its main steps the calculations for checking that the ambiguities are resolvable. In Gröbner basis theory this is even more central, with various completion algorithms being driven by these ambiguity resolution calculations (and adding the twist of modifying the set of reductions whenever it is found that an ambiguity fails to resolve). From the theoretical foundations these algorithms above all else need some criteria for discarding ambiguities that don't need to be checked, as there typically are infinitely many triplets (t, μ, u) which qualify as ambiguities under Definition 5.9. In order to formulate such criteria one needs some extra structure however, and one that performs very well is to have a family of advanceable maps.

What makes advanceable maps useful for structuring a set of ambiguities is primarily that one ambiguity can be the image of another ambiguity; indeed, for e.g. the family (3.8) of advanceable maps *every* image of an ambiguity is another ambiguity. Such images are however never more than shadows of the original ambiguities, since also the ambiguity resolutions can be transported to the image by the advanceable map. This argument (see Lemma 6.2 for the formal claim) is classically used to prove that it is only necessary to check resolvability for minimal ambiguities (since any non-minimal ambiguity would be a shadow of a smaller ambiguity), but it can also be used to prove that it is sufficient to check one labelling of

an expression, since relabelling maps are often advanceable. A catch is however that even if one can skip checking any particular shadow ambiguity, it does not necessarily follow that all shadow ambiguities can be skipped — the "original" of which an ambiguity is a shadow can itself be a shadow of the shadow (e.g. relabellings are typically invertible). The 'shadow-critical' concept of Definition 6.8 is one way around this catch, even though it in general doesn't discard everything that might be skipped.

A major complication when considering shadow ambiguities is that the original ambiguity will often have a different sort — reside in a different base set — than the shadow that one wishes to resolve. Therefore it is in this section necessary to make the multiplicity in the basic framework $\big(\mathcal{M}, R, \mathcal{Y}, \mathcal{O}, T_1(S)\big)$ explicit, and think in terms of a family of such frameworks. Thus there is a set I (the set of "sorts") which serves as the index set for the family of frameworks, and for each $i \in I$ there is a quintuplet $\big(\mathcal{M}(i), R(i), \mathcal{Y}(i), \mathcal{O}(i), T_1(S)(i)\big)$ where:

- $\mathcal{M}(i)$ is a topological abelian group (and $\overline{\mathcal{M}}(i)$ is its completion).
- $R(i)$ is a set of continuous group endomorphisms on $\mathcal{M}(i)$, and hence on $\overline{\mathcal{M}}(i)$.
- $\mathcal{Y}(i)$ is a spanning subset of $\mathcal{M}(i)$.
- $\mathcal{O}(i) = \big\{B_n(i)\big\}_{n=1}^{\infty}$ is a family of $R(i)$-modules (hence subsets of $\mathcal{M}(i)$) that constitutes a family of fundamental neighbourhoods of 0.
- $T_1(S)(i)$ is a set of continuous group endomorphisms on $\overline{\mathcal{M}}(i)$ that commute with elements of $R(i)$.

As with the five main pieces,

- the particular partial order on $\mathcal{Y}(i)$ will be denoted $P(i)$, but all assumptions on this are explicit, like they were in the previous section.

Other things defined from the basic framework are similarly specialised to a sort i by appending an '(i)' to the symbol; the parenthesis notation for indexing may seem a bit peculiar, but it is traditional for operads, which have inspired much of the multi-sorted extensions to this formalism. Operads have $I = \mathbb{N}$ with the index i being the arity of the elements concerned, so there is one set $\mathrm{Irr}(S)(0)$ of irreducible constants, another set $\mathrm{Irr}(S)(1)$ of irreducible unary operations, yet another set $\mathrm{Irr}(S)(2)$ of irreducible binary operations, etc.; other types of algebraic structures typically require other index sets. What happens with respect to ambiguities is that each $T_1(S)(i)$ has its own ambiguities, but it is frequently the case that they turn out to be shadows of ambiguities in some $T_1(S)(i')$.

The family V of maps that one wants to have advanceable reacts differently to the introduction of several sorts: it acquires two sort indices, since

there is no reason the codomain $\overline{\mathcal{M}}(i)$ should have the same sort as the domain $\overline{\mathcal{M}}(i')$. It is however not until Definition 6.8 that this has to be made explicit; before that it is sufficient to reason about specific advanceable maps that relate to specific ambiguities. An underlying set S of rewrite rules will typically not carry sort indices, as every element of it contributes to every $T_1(S)(i)$.

Definition 6.1. Let $i, i' \in I$ be given. A map $v\colon \overline{\mathcal{M}}(i') \longrightarrow \overline{\mathcal{M}}(i)$ is said to be **advanceable** with respect to $T_1(S)(i')$ and $T_1(S)(i)$ if there for every $t' \in T_1(S)(i')$ and $a \in R^*\mathcal{Y}(i')$ exists some $t \in T(S)(i)$ such that $t\big(v(a)\big) = v\big(t'(a)\big)$. The map v is said to be **absolutely advanceable** with respect to $T(S)(i')$ and $T(S)(i)$ if there for every $t' \in T(S)(i')$ exists some $t \in T(S)(i)$ such that $v \circ t' = t \circ v$.

An ambiguity (t, μ, u) of $T_1(S)(i)$ is said to be a **shadow** of the ambiguity (t', μ', u') of $T_1(S)(i')$ if there exists an advanceable continuous homomorphism $v\colon \overline{\mathcal{M}}(i') \longrightarrow \overline{\mathcal{M}}(i)$ such that $\mu = v(\mu')$, $t(\mu) = v\big(t'(\mu')\big)$, and $u(\mu) = v\big(u'(\mu')\big)$. The ambiguity (t, μ, u) is said to be an **absolute shadow** if v is absolutely advanceable.

In the classical case of Bergman's diamond lemma, there is only one sort — preferably denoted 1 for consistency with the operad generalisation — and hence the indices could be dropped. Not dropping indices, one would have $\mathcal{M}(1) = \mathcal{R}\langle X\rangle$, $\mathcal{Y}(1) = X^*$, and $T_1(S)(1)$ being the set of all maps $t_{\nu_1 s \nu_2}$ on the form (3.5); an ambiguity is thus some $(t_{\lambda_1 s_1 \nu_1}, \mu, t_{\lambda_2 s_2 \nu_2})$ where $\lambda_1 \mu_{s_1} \nu_1 = \mu = \lambda_2 \mu_{s_2} \nu_2$. However, if λ_1 and λ_2 have some common prefix κ (i.e., $\lambda_1 = \kappa\lambda_1'$ and $\lambda_2 = \kappa\lambda_2'$ for some $\kappa, \lambda_1', \lambda_2' \in X^*$) and/or ν_1 and ν_2 have some common suffix ρ (i.e., $\nu_1 = \nu_1'\rho$ and $\nu_2 = \nu_2'\rho$ for some $\nu_1', \nu_2', \rho \in X^*$) then for $\mu' = \lambda_1' \mu_{s_1} \nu_1' = \lambda_2' \mu_{s_2} \nu_2'$ one finds that $(t_{\lambda_1 s_1 \nu_1}, \mu, t_{\lambda_2 s_2 \nu_2})$ is a shadow under the absolutely advanceable map $v(b) = \kappa b \rho$ of the ambiguity $(t_{\lambda_1' s_1 \nu_1'}, \mu', t_{\lambda_2' s_2 \nu_2'})$. From the unique factorisation in the free monoid X^*, it follows that the only ambiguities that are not such shadows are those where at least one of λ_1 and λ_2, and at least one of ν_1 and ν_2, are equal to the identity 1. This, with the help of the next two lemmas, cuts down the number of ambiguities that explicitly need to be resolved quite considerably.

Lemma 6.2. *If an ambiguity is resolvable then all its absolute shadows are resolvable as well.*

Proof. Let the indices $i, i' \in I$ be given. Let (t_1', μ', t_2') be a resolvable ambiguity of $T_1(S)(i')$. Let (t_1, μ, t_2) be an arbitrary ambiguity of $T_1(S)(i)$ that is an absolute shadow of (t_1', μ', t_2'), and let $v\colon \overline{\mathcal{M}}(i') \longrightarrow \overline{\mathcal{M}}(i)$ be the absolutely advanceable continuous homomorphism which links the two ambiguities. Let $\varepsilon \in \widehat{\mathcal{O}}(i)$ be arbitrary, and let $\delta \in \widehat{\mathcal{O}}(i')$ be such that $v(\delta) \subseteq \varepsilon$. Since (t_1', μ', t_2') is resolvable there exists $t_3', t_4' \in T(S)(i')$ such that $t_3'\big(t_1'(\mu')\big) - t_4'\big(t_2'(\mu')\big) \in \delta$. By absolute advanceability of v there exists

$t_3, t_4 \in T(S)(i)$ such that $t_3 \circ v = v \circ t'_3$ and $t_4 \circ v = v \circ t'_4$. Then

$$\begin{aligned} t_3\big(t_1(\mu)\big) - t_4\big(t_2(\mu)\big) &= (t_3 \circ t_1)\big(v(\mu')\big) - (t_4 \circ t_2)\big(v(\mu')\big) = \\ &= (v \circ t'_3 \circ t'_1)(\mu') - (v \circ t'_4 \circ t'_2)(\mu) = \\ &= v\big((t'_3 \circ t'_1)(\mu') - (t'_4 \circ t'_2)(\mu')\big) \in v(\delta) \subseteq \varepsilon, \end{aligned}$$

and since ε was arbitrary it follows that (t_1, μ, t_2) is resolvable. □

In this lemma, it would not have been sufficient to assume conditional advanceability, and it is instructive to consider why. Suppose $t_1(\mu')$ is not a single element of $R^*\mathcal{Y}(i')$, but is instead the sum $\lambda + \nu$ of two different elements of $\mathcal{Y}(i')$. Suppose further that $t'_3 \in T_1(S)(i')$, because the extension to non-simple reductions in absolute advanceability is not the main issue. If v is advanceable then there certainly exist $t_{3\lambda}, t_{3\nu} \in T(S)(i)$ such that $t_{3\lambda}\big(v(\lambda)\big) = v\big(t'_3(\lambda)\big)$ and $t_{3\nu}\big(v(\nu)\big) = v\big(t'_3(\nu)\big)$, but there is no guarantee that there is some $t_3 \in T(S)(i)$ such that $t_3\big(v(\lambda+\nu)\big) = v\big(t'_3(\lambda+\nu)\big)$. In very many cases it would probably turn out that something like $t_{3\lambda} \circ t_{3\nu}$ acts exactly as the t_3 one needs, because good choices of simple reductions tend to act trivially on large subsets of $\overline{\mathcal{M}}(i)$, but since it cannot in general be assumed that $t_{3\nu}$ acts trivially on $v(\lambda)$, that composition will sometimes fail.

The idea to first reduce one term, and then the next, is basically good but requires some kind of book-keeping device to work. Provided that simple reductions are of the $t_{\mu\mapsto a}$ kind (i.e., each only acts nontrivially on one element of $\mathcal{Y}$), a possibility would be to use the order on $\mathcal{Y}(i)$ and start with the smallest terms, but a more powerful solution is to go for relative resolvability instead; this provides for reducing different terms independently of each other. The small price one has to pay is a condition on how the advanceable map behaves with respect to the partial orderings.

Lemma 6.3. *Assume the ambiguity (t, μ, u) of $T_1(S)(i)$ is a shadow of the ambiguity (t', μ', u') of $T_1(S)(i')$, that the corresponding advanceable map $v\colon \overline{\mathcal{M}}(i') \longrightarrow \overline{\mathcal{M}}(i)$ satisfies*

$$v\Big(\mathrm{DSM}\big(\mu', P(i')\big)\Big) \subseteq \mathrm{DSM}\big(\mu, P(i)\big), \tag{6.1}$$

and that $T(S)(i)$ is compatible with the partial order $P(i)$. If (t', μ', u') is resolvable relative to $P(i')$, then (t, μ, u) is resolvable relative to $P(i)$.

More generally, an advanceable continuous homomorphism $v\colon \overline{\mathcal{M}}(i') \longrightarrow \overline{\mathcal{M}}(i)$ satisfying (6.1) *also maps $\mathrm{DIS}\big(\mu', P(i'), S\big)$ into $\mathrm{DIS}\big(\mu, P(i), S\big)$ if $T_1(S)(i)$ is compatible with the partial order $P(i)$.*

Proof. Let $D = \mathrm{DIS}\big(\mu, P(i), S\big)$ and $D' = \mathrm{DIS}\big(\mu', P(i'), S\big)$. That (t, μ, u) is resolvable relative to $P(i)$ is by definition that $t(\mu) - u(\mu) \in D$, and since $t(\mu) - u(\mu) = v\big(t'(\mu') - u'(\mu')\big)$ where $t'(\mu') - u'(\mu') \in D'$, the first claim follows from the second: that $v(D') \subseteq D$.

Let $a \in D'$ be given. Since D is topologically closed, it follows that $v(a) \in D$ if it can be shown that $v(a) \in D+\varepsilon$ for every $\varepsilon \in \widehat{\mathcal{O}}(i)$. Let $\varepsilon \in \widehat{\mathcal{O}}(i)$ be arbitrary. Let $\delta \in \widehat{\mathcal{O}}(i')$ be such that $v(\delta) \subseteq \varepsilon$. Let $\{\nu_j\}_{j=1}^m \subseteq \mathcal{Y}(i')$, $\{r_j\}_{j=1}^m \subseteq \pm R^*(i')$, and $\{t'_j\}_{j=1}^m \subseteq T_1(S)(i')$ be such that

$$\sum_{j=1}^{m} r_j\big(\nu_j - t'_j(\nu_j)\big) \in a + \delta$$

and $\nu_j < \mu'$ in $P(i')$ for $j = 1, \ldots, m$. Let $b_j = r_j(\nu_j)$ for $j = 1, \ldots, m$ and let $b = \sum_{j=1}^m \big(b_j - t'_j(b_j)\big)$.

By advanceability of v there exist reductions $\{t_j\}_{j=1}^m \subseteq T(S)(i)$ such that $(t_j \circ v)(b_j) = (v \circ t'_j)(b_j)$ for $j = 1, \ldots, m$. Since $b_j \in \mathrm{DSM}\big(\mu', P(i')\big)$ it follows that $v(b_j) \in \mathrm{DSM}\big(\mu, P(i)\big)$ for $j = 1, \ldots, m$, and hence

$$v(b) = v\Big(\sum_{j=1}^{m}\Big(b_j - t'_j(b_j)\Big)\Big) = \sum_{j=1}^{m}\Big(v(b_j) - t_j\big(v(b_j)\big)\Big) \in D$$

by Lemma 5.10. Furthermore $v(a) - v(b) = v(a - b) \in v(\delta) \subseteq \varepsilon$, thus $v(a) \in D$ by the arbitrariness of ε. □

The following concepts are useful when one seeks to prove that an advanceable map (or family of advanceable maps) satisfies (6.1).

Definition 6.4. A map $v\colon \overline{\mathcal{M}}(i') \longrightarrow \overline{\mathcal{M}}(i)$ is said to **correlate** $P(i')$ to $P(i)$ if

$$v(\mu) \in \mathcal{Y}(i) \quad \Longrightarrow \quad v\Big(\mathrm{DSM}\big(\mu, P(i')\big)\Big) \subseteq \mathrm{DSM}\big(v(\mu), P(i)\big) \tag{6.2}$$

for all $\mu \in \mathcal{Y}(i')$. A map $v\colon \mathcal{Y}(i') \longrightarrow \mathcal{Y}(i)$ is said to be **monotone** with respect to $P(i')$ and $P(i)$ if

$$\nu \leqslant \mu \text{ in } P(i') \quad \Longrightarrow \quad v(\nu) \leqslant v(\mu) \text{ in } P(i) \tag{6.3}$$

for all $\mu, \nu \in \mathcal{Y}(i')$. The map v is **strictly monotone** if

$$\nu < \mu \text{ in } P(i') \quad \Longrightarrow \quad v(\nu) < v(\mu) \text{ in } P(i) \tag{6.4}$$

for all $\mu, \nu \in \mathcal{Y}(i')$.

It needs to be pointed out that there, particularly for functions with domain and codomain in $\mathbb{R}$, exists a conflicting terminology which calls the property defined in (6.3) 'increasing' and instead defines 'monotone' as 'increasing or decreasing'; preferences vary. When 'monotone' as here means "preserves inequalities" then the corresponding name for "reverses inequalities" is *antitone*. Yet another name that might be seen for a map having properties like these is that it is 'compatible' with the order, but here

Definition 5.3 has already given that name to a different relation between maps and binary relations.

The applied concept in this trio is that of a map which correlates $P(i')$ to $P(i)$: it covers the condition (6.1) of Lemma 6.3 and it blends nicely with Construction 7.2 in that it reduces compatibility of simple reductions made from a rule (μ, a) to the matter of whether $a \in \mathrm{DSM}(\mu, P)$. On the other hand, it is usually monotonicity that is the goal when one constructs the relations $\{P(i)\}_{i\in I}$, so a small lemma bridging the gap may be in order.

Lemma 6.5. *Let $i, i' \in I$ be sorts. Let $v\colon \overline{\mathcal{M}}(i') \longrightarrow \overline{\mathcal{M}}(i)$ be a continuous homomorphism such that $v\big(r(\mu)\big) \in \mathrm{Cspan}\big(\{v(\mu)\}\big)$ for all $\mu \in \mathcal{Y}(i')$ and $r \in R^*(i')$. Let $P(i)$ be a binary relation on $\mathcal{Y}(i)$ and let $P(i')$ be a binary relation on $\mathcal{Y}(i')$. If $v(\nu) \in \mathrm{DSM}\big(v(\mu), P(i)\big)$ for all $\mu, \nu \in \mathcal{Y}$ such that $\nu < \mu$ in $P(i')$ and $v(\mu) \in \mathcal{Y}$, then v correlates $P(i')$ to $P(i)$.*

Proof. Let $\mu \in \mathcal{Y}(i')$ such that $v(\mu) \in \mathcal{Y}$ be given. Let $D = \mathrm{DSM}\big(v(\mu), P(i)\big)$. If $\nu \in \mathcal{Y}(i')$ satisfies $\nu < \mu$ in $P(i')$ then by assumption $v(\nu) \in D$ and hence $\mathrm{Cspan}\big(\{v(\nu)\}\big) \subseteq D$, which implies $v\big(r(\nu)\big) \in D$ for all $r \in R^*(i')$. Since v is a continuous homomorphism and D is a topologically closed group, it now follows that

$$v\Big(\mathrm{Cspan}\big(\{\, \nu \in \mathcal{Y}(i') \,\big|\, \nu < \mu \text{ in } P(i') \,\}\big)\Big) \subseteq D,$$

i.e., v satisfies the condition at μ for correlating $P(i')$ to $P(i)$. □

Remark. The meaning of (6.4) if v ranges over all maps in the family (3.8), as would be the setting for Bergman's diamond lemma, is that

$$\nu < \mu \text{ in } P(1) \quad \Longrightarrow \quad \lambda\nu\rho < \lambda\mu\rho \text{ in } P(1) \qquad \text{for all } \mu, \nu, \lambda, \rho \in X^*, \tag{6.5}$$

i.e., $P(1)$ must be a (strictly compatible) monoid partial order. The need for this very classical condition can thus in the generic theory be found only in the resolvability of ambiguities! It would be possible to apply the *generic* diamond lemma with a partial order that violates (6.5), but for that one would then pay the price that ambiguity resolution gets more complicated. Furthermore compatibility of reductions interacts with advanceability so that one anyway comes pretty close to needing correlation; in practice the choice one has is one of how many advanceable maps one will have rather than whether these will correlate the partial orders.

The condition that advanceable maps should be continuous can similarly be regarded as a condition on how the multiplication operation on $\mathcal{M}(1) = \mathcal{R}\langle X\rangle$ should relate to the topology, and if for example (2.6) holds then continuity of maps on the form (3.8) can also be reduced to a condition on multiplication of monomials, but this point of view is not as striking as it is for the order on $\mathcal{Y}$, since continuity to the average mathematician is more of an everyday condition.

In most classical cases, Lemma 6.5 would be applied to maps v which map $\mathcal{Y}(i')$ into $\mathcal{Y}(i)$, in which case correlation implies strict monotonicity. Some algebraic structures will however give rise to "degenerate"maps which cannot be strictly monotone on the whole of $\mathcal{Y}(i')$, and at least in the notable case of path algebras (where the product of two monomials can be zero, see Subsection 8.3) the additional precondition that $v(\mu) \in \mathcal{Y}(i)$ provides a convenient loophole to avoid getting caught by this technicality.

The final nontrivial condition in Lemma 6.5 is that v should map elements on the form $r(\mu)$ into $\mathrm{Cspan}\big(\{v(\mu)\}\big)$. The most common reason this condition would be fulfilled is that $\mathcal{M}(i')$ and $\mathcal{M}(i)$ are both $\mathcal{R}$-modules for some ring $\mathcal{R}$ such that the advanceable map $v\colon \mathcal{M}(i') \longrightarrow \mathcal{M}(i)$ is $\mathcal{R}$-linear, while $R(i')$ and $R(i)$ are the sets of actions of elements of $\mathcal{R}$ on $\mathcal{M}(i')$ and $\mathcal{M}(i)$ respectively. Note, however, that the general framework makes no assumption that maps in $R(i')$ should have counterparts in $R(i)$, or even that there should be a corresponding endomorphism on $\overline{\mathcal{M}}(i)$. This is a reason why advanceability is a condition on how arbitrary elements of $R^*\mathcal{Y}$ are treated; for suitably linear maps it is sufficient to check advanceability on elements of $\mathcal{Y}$.

Whether it could be useful to have $R(i')$ and $R(i)$ generate nonisomorphic rings of endomorphisms on $\mathcal{M}(i')$ and $\mathcal{M}(i)$ respectively remains to be seen, but not requiring $\mathcal{R}$-linearity turns out to be advantageous for the related concept of *biadvanceability*. (Shadow ambiguities are not the only ones that are traditionally discarded; equally important are ambiguities where the parts are disjoint. Biadvanceability provides a way to give an abstract definition of this.) Recall that an $\mathcal{R}$-bilinear map w satisfies $w(ra, b) = rw(a, b) = w(a, rb)$, from which follows $rsw(a, b) = rw(a, sb) = w(ra, sb) = sw(ra, b) = srw(a, b)$, for all $r, s \in \mathcal{R}$. If $\mathcal{R}$ is commutative this is only natural, but if $\mathcal{R}$ is a noncommutative ring then it places a rather severe restriction on the range of w: only elements at which all $\mathcal{R}$-module actions commute are allowed! This would often be insufficient for the intended uses of biadvanceable maps.

A practical compromise that is sometimes available is to request some weak form of bilinearity. It might for example be the case that the identity $w(ra, b) = rw(a, b) = w(a, rb)$ only holds for monomial a and b. It could also be the case that the identity is relaxed to $w(ra, b) = r'w(a, b)$ (and similarly for moving out from the second position), where $r' \in \mathcal{R}$ need not be equal to r and may depend on a or b. The condition needed for Lemma 6.5 is weaker still — roughly that $w(ra, b), w(a, rb) \in \mathrm{Cspan}\big(\{w(a, b)\}\big)$ for $a, b \in \mathcal{Y}$; see (6.6) — and should therefore not be a problem to fulfil when necessary.

Definition 6.6. Let $i, i_1, i_2 \in I$ be sorts. A map $w\colon \overline{\mathcal{M}}(i_1) \times \overline{\mathcal{M}}(i_2) \longrightarrow \overline{\mathcal{M}}(i)$ is said to be a **bihomomorphism** if

$$w(a_1 - b_1, a_2 - b_2) = w(a_1, a_2) - w(b_1, a_2) - w(a_1, b_2) + w(b_1, b_2)$$

for all $a_1, b_1 \in \overline{\mathcal{M}}(i_1)$ and $a_2, b_2 \in \overline{\mathcal{M}}(i_2)$. A bihomomorphism $w\colon \overline{\mathcal{M}}(i_1) \times \overline{\mathcal{M}}(i_2) \longrightarrow \overline{\mathcal{M}}(i)$ is said to be **biadvanceable** (with respect to $T(S)(i)$, $T(S)(i_1)$, and $T(S)(i_2)$) if

- there for every $t_1 \in T_1(S)(i_1)$, $b_1 \in R^*\mathcal{Y}(i_1)$, and $b_2 \in R^*\mathcal{Y}(i_2)$ exists some $t \in T(S)(i)$ such that $t\big(w(b_1, b_2)\big) = w\big(t_1(b_1), b_2\big)$, and
- there for every $t_2 \in T(S)(i_2)$, $b_1 \in R^*\mathcal{Y}(i_1)$, and $b_2 \in R^*\mathcal{Y}(i_2)$ exists some $t \in T(S)(i)$ such that $t\big(w(b_1, b_2)\big) = w\big(b_1, t_2(b_2)\big)$.

An ambiguity (t, μ, u) of $T_1(S)(i)$ is said to be a **montage** of the **pieces** $(\lambda, t') \in \mathcal{Y}(i_1) \times T(S)(i_1)$ and $(\nu, u') \in \mathcal{Y}(i_2) \times T(S)(i_2)$ if there exists a continuous biadvanceable map $w\colon \mathcal{M}(i_1) \times \mathcal{M}(i_2) \longrightarrow \mathcal{M}(i)$ such that

$$\begin{aligned} \mu &= w(\lambda, \nu), \\ t\big(w(\lambda, \nu)\big) &= w\big(t'(\lambda), \nu\big) \\ u\big(w(\lambda, \nu)\big) &= w\big(\lambda, u'(\nu)\big) \end{aligned}$$

The map w is called the **composition map** of this montage.

Let V_1 be a set of maps $\overline{\mathcal{M}}(i_1) \longrightarrow \overline{\mathcal{M}}(i)$ and let V_2 be a set of maps $\overline{\mathcal{M}}(i_2) \longrightarrow \overline{\mathcal{M}}(i)$. A biadvanceable map $w\colon \overline{\mathcal{M}}(i_1) \times \overline{\mathcal{M}}(i_2) \longrightarrow \overline{\mathcal{M}}(i)$ is said to be (V_1, V_2)**-biadvanceable** if $w(\cdot, \rho) \in V_1$ for all $\rho \in \mathcal{Y}(i_2)$ and $w(\rho, \cdot) \in V_2$ for all $\rho \in \mathcal{Y}(i_1)$. A (V_1, V_2)**-montage ambiguity** is a montage ambiguity where the composition map is (V_1, V_2)-biadvanceable.

The idea formalised by the montage ambiguity concept is to recognise the situation that the two pieces act on disjoint parts of μ — the pieces are like two small windows to completely different gardens that have been embedded into a large mural provided by the composition map, and the presence of an embedding cannot change the fact that the games that can be played in one garden are quite independent of what happens in the other. In Gröbner basis theory this idea [4] is known as *Buchberger's First Criterion* for eliminating useless critical pairs, although the identification is perhaps not obvious at this stage. The correspondence will be made clear in Corollary 8.5, however.

In the $i = i_1 = i_2 = 1$ case of $\mathcal{M}(1) = \mathcal{R}\langle X\rangle$ and $\mathcal{Y}(1) = X^*$, with all maps of the form $b \mapsto \nu_1 b \nu_2$ for some $\nu_1, \nu_2 \in X^*$ being advanceable, the typical form of a biadvanceable map is $w(a, b) = \nu_1 a \nu_2 b \nu_3$ for some $\nu_1, \nu_2, \nu_3 \in X^*$. Such 'multiplication with fixed extra factors' maps can be used to produce a great variety of biadvanceable maps, and the construction does not require the multiplication operation to be associative, or even binary; pretty much anything that can be composed from multilinear operations on and fixed elements of $\mathcal{Y}$ will probably turn out to be biadvanceable, if simple reductions are constructed by putting every rule in every possible context. The underlying idea for making a biadvanceable map is however to take an element of $\mathcal{Y}$ and cut out two disjoint pieces from it — the biadvanceable map then consists of inserting the two arguments into these

two holes. How such a map may be interpreted depends very much on the underlying algebraic structure, but for the diamond lemma machinery it is sufficient that the biadvanceable maps exist.

Lemma 6.7. *Let (t, μ, u) be an ambiguity of $T_1(S)(i)$ that is a montage of the pieces $(\lambda, t') \in \mathcal{Y}(i_1) \times T_1(S)(i_1)$ and $(\nu, u') \in \mathcal{Y}(i_2) \times T_1(S)(i_2)$. If $T(S)(j)$ is compatible with some partial order $P(j)$ on $\mathcal{Y}(j)$ for all $j \in \{i, i_1, i_2\}$ and the composition map $w\colon \mathcal{M}(i_1) \times \mathcal{M}(i_2) \longrightarrow \mathcal{M}(i)$ of the montage satisfies*

$$w\Big(\mathrm{DSM}\big(\lambda, P(i_1)\big), \nu\Big) \cup w\Big(\lambda, \mathrm{DSM}\big(\nu, P(i_2)\big)\Big) \subseteq \mathrm{DSM}\big(w(\lambda, \nu), P(i)\big) \quad (6.6)$$

then (t, μ, u) is resolvable relative to $P(i)$.

Proof. The problem is to prove that

$$\begin{aligned} t(\mu) - u(\mu) &= w\big(t'(\lambda), \nu\big) - w\big(\lambda, u'(\nu)\big) = \\ &= w\big(t'(\lambda), \nu - u'(\nu)\big) + w\big(t'(\lambda) - \lambda, u'(\nu)\big) \in \mathrm{DIS}\big(\mu, P(i), S\big), \end{aligned}$$

and by symmetry it is sufficient to do the first of $w\big(t'(\lambda), \nu - u'(\nu)\big)$ and $w\big(t'(\lambda) - \lambda, u'(\nu)\big)$, as the other is completely analogous.

Let $\varepsilon \in \widehat{\mathcal{O}}(i)$ be arbitrary. Let $\varepsilon_1 \in \widehat{\mathcal{O}}(i_1)$ be such that $w\big(\varepsilon_1, \nu - u'(\nu)\big) \subseteq \varepsilon$. Since $t'(\lambda) \in \mathrm{DSM}\big(\lambda, P(i_1)\big)$ there exist $\{\lambda_j\}_{j=1}^m \subseteq \mathcal{Y}(i_1)$ and $\{r_j\}_{j=1}^m \subseteq \pm R^*(i_1)$ such that $\lambda_j < \lambda$ in $P(i_1)$ for all $j = 1, \ldots, m$ and $\sum_{j=1}^m r_j(\lambda_j) \in t'(\lambda) + \varepsilon_1$. Let $a_j = r_j(\lambda_j)$ for $j = 1, \ldots, m$. By biadvanceability there exist $\{u_j\}_{j=1}^m \subseteq T(S)(i)$ such that $u_j\big(w(a_j, \nu)\big) = w\big(a_j, u'(\nu)\big)$ for all $j = 1, \ldots, m$. Since $w(a_j, \nu) \in \mathrm{DSM}\big(\mu, P(i)\big)$ by (6.6), these satisfy

$$\begin{aligned} w\Big(\sum_{j=1}^m a_j, \nu - u'(\nu)\Big) &= \sum_{j=1}^m \Big(w(a_j, \nu) - w\big(a_j, u'(\nu)\big)\Big) = \\ &= \sum_{j=1}^m \Big(w(a_j, \nu) - u_j\big(w(a_j, \nu)\big)\Big) \in \\ &\in \sum_{j=1}^m \Big\{\, a - u_j(a) \;\Big|\; a \in \mathrm{DSM}\big(\mu, P(i)\big) \,\Big\} \subseteq \\ &\subseteq \mathrm{DIS}\big(\mu, P(i), S\big) \end{aligned}$$

by Lemma 5.10. Therefore $w\big(t'(\lambda), \nu - u'(\nu)\big) \in \mathrm{DIS}\big(\mu, P(i), S\big) + \varepsilon$, and by the arbitrariness of ε thus $w\big(t'(\lambda), \nu - u'(\nu)\big) \in \mathrm{DIS}\big(\mu, P(i), S\big)$, as claimed. □

The pieces are now in place for a definition of *critical* as in 'critical pair', i.e., "member of a (small) set of ambiguities that together cover all ways in which things can fail to resolve". The definition given is with respect to a

particular family of advanceable maps, since this is how it will typically be applied: when someone considers only *those* advanceable maps, then *these* are the ambiguities that need to be explicitly checked. It is often natural to let the family of advanceable maps be (the set of morphisms in) a category, but there is no technical need for this.

Definition 6.8. Let a family $V = \bigcup_{i,j\in I} V(i,j)$ of maps such that every $v \in V(i,j)$ is an advanceable continuous homomorphism $\overline{\mathcal{M}}(j) \longrightarrow \overline{\mathcal{M}}(i)$ be given.

The family V is said to be a **category** if $V(i,i) \ni \mathrm{id}\colon \overline{\mathcal{M}}(i) \longrightarrow \overline{\mathcal{M}}(i)$ and $v_2 \circ v_1 \in V(i,k)$ for all $v_2 \in V(i,j)$, $v_1 \in V(j,k)$, and $i,j,k \in I$. The family $V = \bigcup_{i,j\in I} V(i,j)$ is the **category generated by** $V_1 = \bigcup_{i,j\in I} V_1(i,j)$ if it is the smallest category that satisfies $V_1(i,j) \subseteq V(i,j)$ for all $i,j \in I$.

An ambiguity (t,μ,u) of $T_1(S)(i)$ is said to be a **V-shadow** of the ambiguity (t',μ',u') of $T_1(S)(i')$ if there exists some $v \in V(i,i')$ such that $\mu = v(\mu')$, $t(\mu) = v\big(t'(\mu')\big)$, and $u(\mu) = v\big(u'(\mu')\big)$. If in addition (t',μ',u') is not a V-shadow of (t,μ,u) then (t,μ,u) is a **proper V-shadow** of (t',μ',u'). An ambiguity of $T_1(S)$ is said to be **V-shadow-minimal** if it is not a proper V-shadow of any ambiguity of $T_1(S)$. An ambiguity of $T_1(S)$ is said to be **V-shadow-critical** if it is not a proper V-shadow of any V-shadow-minimal ambiguity of $T_1(S)$.

An ambiguity (t,μ,u) of $T_1(S)(i)$ is said to be **V-critical** if it is V-shadow-critical and is not a $\big(V(i,i_1),V(i,i_2)\big)$-montage ambiguity for any $i_1,i_2 \in I$.

If V is a category then the V-shadow relation Q_V — formally defined by $(t,\mu,u) \geqslant (t',\mu',u')$ in Q_V iff (t,μ,u) is a V-shadow of (t',μ',u') — is a quasi-order on the set of ambiguities. This point of view is instructive for understanding the definition of V-shadow-critical; (t,μ,u) is a proper V-shadow of (t',μ',u') iff $(t,\mu,u) > (t',\mu',u')$ in Q_V and (t,μ,u) is V-shadow-minimal iff it is Q_V-minimal. A first stab at defining V-shadow-critical would be to use V-shadow-minimal, reasoning that non-minimal ambiguities need not be considered critical as there is always some smaller ambiguity of which they are a shadow, but this fails if Q_V is not DCC; a simple example of a family V for which this might occur is $V = \{D^n\}_{n=0}^{\infty}$, where $D(1) = 0$ and $D(\mathsf{x}^{n+1}) = \mathsf{x}^n$ for all $n \in \mathbb{N}$. By only discarding those ambiguities which are proper shadows of a minimal ambiguity, one arrives at a concept which is as strong as minimality in the nice cases but is sufficient also in the strange cases.

Theorem 6.9. *Let $V = \bigcup_{i,j\in I} V(i,j)$ be a family of maps such that every $v \in V(i,j)$ is an advanceable continuous homomorphism $\overline{\mathcal{M}}(j) \longrightarrow \overline{\mathcal{M}}(i)$. For each $i \in I$, let $P(i)$ be a partial order on $\mathcal{Y}(i)$ with which $T_1(S)(i)$ is compatible. If every $v \in V(i,j)$, for all $i,j \in I$, correlates $P(j)$ to $P(i)$ then the following claims are equivalent:*

(a′) *Every ambiguity of* $T_1(S)(i)$ *is resolvable relative to* $P(i)$*, for all* $i \in I$*.*

(a″) *Every* V*-critical ambiguity of* $T_1(S)(i)$ *is resolvable relative to* $P(i)$*, for all* $i \in I$*.*

Proof. Since the V-critical ambiguities of (a″) are included among the ambiguities of (a′), all that needs to be shown is that the non-V-critical ambiguities are resolvable relative to P whenever the V-critical ambiguities are so resolvable. Hence assume (a″).

If an ambiguity (t, μ, u) of $T_1(S)(i)$ is not V-shadow-critical, then by definition there exists some V-shadow-minimal ambiguity, say (t', μ', u') of $T_1(S)(i')$, of which (t, μ, u) is a proper V-shadow. Since (t', μ', u') is minimal it is not a proper V-shadow of any ambiguity, and hence (t', μ', u') is V-critical. By (a″), (t', μ', u') is resolvable relative to $P(i')$, which by Lemma 6.3 implies that (t, μ, u) is resolvable relative to $P(i)$, as claimed.

If an ambiguity (t, μ, u) of $T_1(S)(i)$ is a $\big(V(i, i_1), V(i, i_2)\big)$-montage ambiguity for some $i_1, i_2 \in I$ then it is resolvable relative to $P(i)$ by Lemma 6.7; (6.6) holds because $V(i, i_1)$ is a set of maps correlating $P(i_1)$ to $P(i)$, $V(i, i_2)$ is a set of maps correlating $P(i_2)$ to $P(i)$, and the composition map w of the ambiguity (t, μ, u) is $\big(V(i, i_1), V(i, i_2)\big)$-biadvanceable. □

It should be observed that the set of V-critical ambiguities is not always the smallest set of ambiguities with which one can make do; if (t, μ, u) is V-critical then every (t', μ', u') such that $(t, \mu, u) \sim (t', \mu', u')$ in Q_V is V-critical as well, even though it is clearly sufficient to check one ambiguity in each Q_V-equivalence class. In actual calculations this often corresponds to being able to pick one labelling of an ambiguity and resolve it in that context, instead of having to write a resolution proof for arbitrary labellings.

Another labour-saving trick which goes beyond the definition of V-critical ambiguity is *Buchberger's Second Criterion*, which in its raw form is simply the observation that if three simple reductions t_1, t_2, and t_3 act nontrivially on the same μ, then relative resolvability of two of the resulting ambiguities (t_1, μ, t_2), (t_1, μ, t_3), and (t_2, μ, t_3) implies the same for the third. Under mild extra assumptions on V, this criterion can be given the more traditional form that (t_1, μ, t_2) can be skipped if there exists some simple reduction t_3 such that (t_1, μ, t_3) and (t_2, μ, t_3) are both non-shadow-critical, since the latter two can then be assumed relatively resolvable on account of being shadows of other ambiguities. This criterion is of practical interest because it is often far less work to perform an explicit search for a matching t_3 than it is to explicitly resolve (t_1, μ, t_2), but it is not as theoretically important as the recognition of montage ambiguities (Buchberger's first criterion).

Example 6.10. Let $I = \{1\}$ and drop sort indices. Let $\mathcal{M} = \mathcal{R}\langle X\rangle$, $\mathcal{Y} = X^*$, $R = \mathcal{R}$, the topology be discrete, $V = \{b \mapsto \nu_1 b \nu_2\}_{\nu_1, \nu_2 \in X^*}$, and

the simple reductions be defined as in Corollary 3.9 (or equivalently Construction 7.2); this is the setting for Bergman's diamond lemma. Which are then the V-critical ambiguities?

Using the notation of (3.5), an ambiguity has the form $(t_{\lambda_1 s_1 \nu_1}, \mu, t_{\lambda_2 s_2 \nu_2})$ where $\lambda_1 \mu_{s_1} \nu_1 = \mu = \lambda_2 \mu_{s_2} \nu_2$. By unique factorisation of μ in X^*, λ_1 is a prefix (left divisor) of λ_2 or vice versa — hence there exist $\kappa \in \{\lambda_1, \lambda_2\}$ and $\lambda_1', \lambda_2' \in X^*$ such that $\lambda_1 = \kappa \lambda_1'$ and $\lambda_2 = \kappa \lambda_2'$. Similarly ν_1 is a suffix (right divisor) of ν_2 or vice versa, whence there exist $\rho \in \{\nu_1, \nu_2\}$ and $\nu_1', \nu_2' \in X^*$ such that $\nu_1 = \nu_1' \rho$ and $\nu_2 = \nu_2' \rho$. It follows that $(t_{\lambda_1 s_1 \nu_1}, \mu, t_{\lambda_2 s_2 \nu_2})$ is a shadow under $v(b) = \kappa b \rho$ of $(t_{\lambda_1' s_1 \nu_1'}, \mu', t_{\lambda_2' s_2 \nu_2'})$ where $\mu' = \lambda_1' \mu_{s_1} \nu_1' = \lambda_2' \mu_{s_2} \nu_2'$, and this shadow is proper unless $\kappa = \rho = 1$. Conversely any ambiguity $(t_{\lambda_1 s_1 \nu_1}, \mu, t_{\lambda_2 s_2 \nu_2})$ which is a proper V-shadow must have $\lambda_1, \lambda_2 \neq 1$ or $\nu_1, \nu_2 \neq 1$, so it follows that the constructed $(t_{\lambda_1' s_1 \nu_1'}, \mu', t_{\lambda_2' s_2 \nu_2'})$ is V-shadow-minimal. Hence a V-shadow-critical ambiguity $(t_{\lambda_1 s_1 \nu_1}, \mu, t_{\lambda_2 s_2 \nu_2})$ has $1 \in \{\lambda_1, \lambda_2\}$ and $1 \in \{\nu_1, \nu_2\}$. Without loss of generality it may be assumed that $\lambda_1 = 1$.

If the ambiguity $(t_{1 s_1 \nu_1}, \mu, t_{\lambda_2 s_2 \nu_2})$ is such that μ_{s_1} is a prefix of λ_2, say $\lambda_2 = \mu_{s_1} \tau$, then conversely $\mu_{s_2} \nu_2$ is a suffix of $\nu_1 = \tau \mu_{s_2} \nu_2$, and consequently $(t_{1 s_1 \nu_1}, \mu, t_{\lambda_2 s_2 \nu_2})$ is a montage with the composition map $w(b_1, b_2) = b_1 \tau b_2$ of the pieces $(\mu_{s_1}, t_{1 s_1 1})$ and $(\mu_{s_2} \nu_2, t_{1 s_2 \nu_2})$. Hence a V-critical ambiguity $(t_{1 s_1 \nu_1}, \mu, t_{\lambda_2 s_2 \nu_2})$ rather has $\mu = \lambda_2 \tau \nu$, where either (if $\nu_1 = 1$: an *inclusion* ambiguity) $\tau = \mu_{s_2}$, $\nu = \nu_2$, and $\mu_{s_1} = \lambda_2 \tau \nu$ or (if $\nu_1 \neq 1$: an *overlap* ambiguity) $\mu_{s_1} = \lambda_2 \tau$, $\mu_{s_2} = \tau \nu$, and $\nu = \nu_1$. Either way, the ambiguity is uniquely identified by the quintuplet $(s_1, s_2, \lambda_2, \tau, \nu)$, which (as it happens) is the definition of ambiguity that was used in [2]. This has shown that all V-critical ambiguities are among those specified by Bergman and consequently Bergman's diamond lemma follows from combining Theorems 5.11 and 6.9.

It may also be observed that if the set S of rules is finite then the set of V-critical ambiguities is finite as well; for any given pair $(s_1, s_2) \in S^2$, a non-montage ambiguity on the form $(t_{1 s_1 \nu_1}, \mu, t_{\lambda_2 s_2 \nu_2})$ where at least one of ν_1 and ν_2 is equal to 1 must satisfy $\deg \lambda_2 < \deg \mu_{s_1}$, and hence the number of V-critical ambiguities on this form can be at most $\deg \mu_{s_1}$. This gives the overall bound $|S|^2 \max_{s \in S} \deg \mu_s$, or sharper $|S| \sum_{s \in S} \deg \mu_s$, for the number of V-critical ambiguities, but the actual number is often much lower.

Note how the finiteness of the set of V-critical ambiguities whenever S is finite requires that the montage ambiguities are discarded. This is not the case in the commutative counterpart $\mathcal{M} = \mathcal{R}[X]$, since in that case all montage ambiguities are shadows of the one with $w(a, b) = ab$. This is probably the reason that this important principle in the commutative theory is merely known as the "first criterion".

As was parenthetically remarked in the example, Bergman distinguishes between *inclusion* and *overlap* ambiguities, where the former have the prop-

erty that μ_{s_2} divides (is a subword of) μ_{s_1}. Since the property of being a divisor can be expressed in terms of advanceable maps, these classes may be defined also in the more abstract setting.

Definition 6.11. An ambiguity (t, μ, u) of $T_1(S)(i)$ is said to be an **inclusion ambiguity**, where t is called the **inner reduction** and u is called the **outer reduction**, if there for every advanceable map $v\colon \overline{\mathcal{M}}(i') \longrightarrow \overline{\mathcal{M}}(i)$ and $(\mu', u') \in \mathcal{Y}(i') \times T_1(S)(i')$ such that $\mu = v(\mu')$ and $u(\mu) = v\big(u'(\mu')\big)$ exists some $t' \in T_1(S)(i')$ such that $t(\mu) = v\big(t'(\mu')\big)$. The inclusion is said to be **proper** if only one of the reductions fit the definition for being the inner reduction.

An ambiguity is said to be an **overlap ambiguity** if it is neither a montage ambiguity nor an inclusion ambiguity.

For handmade sets of simple reductions $T_1(S)$ (or ditto sets of rules S from which they are made), inclusion ambiguities are rare, because they typically mean the outer reduction is redundant and can be dropped without changing $\mathrm{Irr}(S)$ (by Theorem 5.6), $\mathrm{Red}(S)$, or t^S (by Lemma 4.11). The situation is a bit different in sets of reductions that are automatically generated by some completion procedure, since it is very common that special cases of a rule are derived before (and even used in deriving) the more general rule that one may find in the literature. Relying on Lemma 4.11 for simplifying the set of reductions would require keeping all reductions until a complete set is found and only then drop those which are redundant, but it is usually more practical to drop them as soon as the inclusion is discovered. The next theorem gives conditions for this.

Theorem 6.12. *For every $i \in I$, let $T_1(S')(i) \subseteq T_1(S)(i)$ and a partial order $P(i)$ on $\mathcal{Y}(i)$ with which $T_1(S)(i)$ is compatible be given. Assume there is some $i_0 \in I$ and $t_0 \in T_1(S)(i_0) \setminus T_1(S')(i_0)$ such that there for every $i \in I$, $t \in T_1(S)(i) \setminus T_1(S')(i)$, and $\mu \in \mathcal{Y}(i)$ on which t acts nontrivially exists a continuous homomorphism $v\colon \overline{\mathcal{M}}(i_0) \longrightarrow \overline{\mathcal{M}}(i)$ which is advanceable with respect to $T_1(S')(i_0)$ and $T_1(S')(i)$ and also some $\mu_0 \in \mathcal{Y}(i_0)$ such that $v(\mu_0) = \mu$, $t(\mu) = v\big(t_0(\mu_0)\big)$, and $v\big(\mathrm{DSM}(\mu_0, P(i_0))\big) \subseteq \mathrm{DSM}(\mu, P(i))$.*

If for all $i \in I$ all ambiguities of $T_1(S')(i)$ are resolvable with respect to $P(i)$ and there for every $\mu_0 \in \mathcal{Y}(i_0)$ on which t_0 acts nontrivially exists some $u_0 \in T_1(S')(i_0)$ such that $t_0(\mu_0) - u_0(\mu_0) \in \mathrm{DIS}\big(\mu_0, P(i_0), S'\big)$, then for all $i \in I$ all ambiguities of $T_1(S)(i)$ are resolvable with respect to $P(i)$ and $\mathcal{I}(S')(i) = \mathcal{I}(S)(i)$.

Proof. Let $i \in I$ be arbitrary. Let (t, μ, u) be an arbitrary ambiguity of $T_1(S)(i)$. If $t, u \in T_1(S')(i)$ then $t(\mu) - u(\mu) \in \mathrm{DIS}\big(\mu, P(i), S'\big) \subseteq \mathrm{DIS}\big(\mu, P(i), S\big)$ by assumption.

If $t \in T_1(S)(i) \setminus T_1(S')(i)$ and $u \in T_1(S')(i)$ then by assumption there exists some $\mu_0 \in \mathcal{Y}(i_0)$ and an advanceable continuous homomorphism

$v\colon \overline{\mathcal{M}}(i_0) \longrightarrow \overline{\mathcal{M}}(i)$ such that $v\big(\mathrm{DSM}(\mu_0, P(i_0))\big) \subseteq \mathrm{DSM}(\mu, P(i))$, $\mu = v(\mu_0)$, and $t(\mu) = v\big(t_0(\mu_0)\big)$. Since $v\big(t_0(\mu_0)\big) = t(\mu) \neq \mu = v(\mu_0)$ it follows that t_0 acts nontrivially on μ_0, and hence there exists some $u_0 \in T_1(S')(i_0)$ such that $t_0(\mu_0) \equiv u_0(\mu_0) \pmod{S' < \mu_0 \text{ in } P(i_0)}$. By the second claim of Lemma 6.3, $v\big(t_0(\mu_0)\big) \equiv v\big(u_0(\mu_0)\big) \pmod{S' < \mu \text{ in } P(i)}$, and by advanceability there exists some $u_1 \in T(S')(i)$ such that $u_1(\mu) = v\big(u_0(\mu_0)\big)$; in other words $t(\mu) \equiv u_1(\mu) \pmod{S' < \mu \text{ in } P(i)}$.

It need not be the case that $u_1 = u$, or even that $u_1 \in T_1(S')(i)$, but typically $u_1(\mu) \neq \mu$ and then there exist $u_{1a} \in T_1(S')(i)$ and $u_{1b} \in T(S')(i)$ such that $u_1(\mu) = u_{1b}\big(u_{1a}(\mu)\big)$ and $u_{1a}(\mu) \neq \mu$. In this case, (u, μ, u_{1a}) is an ambiguity of $T_1(S')(i)$ and $u(\mu) \equiv u_{1a}(\mu) \pmod{S' < \mu \text{ in } P(i)}$ since it is resolvable. Furthermore $u_{1a}(\mu) \equiv u_1(\mu) \pmod{S' < \mu \text{ in } P(i)}$ by Lemma 5.10, and it follows that (t, μ, u) is resolvable relative to $P(i)$. In the degenerate case that $u_1(\mu) = \mu$, one has the curious situation that $\mu = v\big(u_0(\mu_0)\big) \in v\big(\mathrm{DSM}(\mu_0, P(i_0))\big) \subseteq \mathrm{DSM}\big(\mu, P(i)\big)$; there must be an alternative expression for μ as a linear combination of strictly smaller elements of $\mathcal{Y}(i)$. Hence $\mu - u(\mu) \in \mathrm{DIS}\big(\mu, P(i), S'\big)$ by definition and therefore $t(\mu) \equiv u_1(\mu) = \mu \equiv u(\mu) \pmod{S' < \mu \text{ in } P(i)}$.

The case $t \in T_1(S')(i)$ and $u \in T_1(S)(i) \setminus T_1(S')(i)$ is handled similarly. The case $t, u \in T_1(S)(i) \setminus T_1(S')(i)$ is handled by combining the two previous cases — having two advanceable maps $v, v'\colon \overline{\mathcal{M}}(i_0) \longrightarrow \overline{\mathcal{M}}(i)$ and $\mu_0, \mu_0' \in \mathcal{Y}(i_0)$ such that $v(\mu_0) = \mu = v'(\mu_0')$, $t(\mu) = v\big(t_0(\mu_0)\big)$, and $u(\mu) = v'\big(t_0(\mu_0')\big)$.

As for the claim that $\mathcal{I}(S')(i) = \mathcal{I}(S)(i)$, it follows from Lemma 3.7 that

$$\begin{aligned}\mathcal{I}(S')(i) &= \mathrm{Cspan}\Big(\big\{\, \mu - t(\mu) \,\big|\, \mu \in \mathcal{Y}(i), t \in T_1(S')(i) \,\big\}\Big) \subseteq \\ &\subseteq \mathrm{Cspan}\Big(\big\{\, \mu - t(\mu) \,\big|\, \mu \in \mathcal{Y}(i), t \in T_1(S)(i) \,\big\}\Big) = \mathcal{I}(S)(i).\end{aligned}$$

Furthermore, if $t \in T_1(S)(i) \setminus T_1(S')(i)$ and $\mu \in \mathcal{Y}(i)$ can give a nonzero contribution to the second Cspan, i.e., if they are such that $\mu - t(\mu) \neq 0$, then by assumption there exist $\mu_0 \in \mathcal{Y}(i_0)$ and an advanceable continuous homomorphism $v\colon \overline{\mathcal{M}}(i_0) \longrightarrow \overline{\mathcal{M}}(i)$ such that $\mu = v(\mu_0)$, $t(\mu) = v\big(t_0(\mu_0)\big)$, and $v\big(\mathrm{DSM}(\mu_0, P(i_0))\big) \subseteq \mathrm{DSM}(\mu, P(i))$. There also exists some $u_0 \in T_1(S')(i_0)$ such that $t_0(\mu_0) \equiv u_0(\mu_0) \pmod{S' < \mu_0 \text{ in } P(i_0)}$ and some $u \in T(S')(i)$ such that $u\big(v(\mu_0)\big) = v\big(u_0(\mu_0)\big)$. Hence

$$\begin{aligned}\mu - t(\mu) = v\big(\mu_0 - t_0(\mu_0)\big) = v\big(\mu_0 - u_0(\mu_0)\big) + v\big(u_0(\mu_0) - t_0(\mu_0)\big) \in \\ \in \big(\mu - u(\mu)\big) + v\Big(\mathrm{DIS}\big(\mu_0, P(i_0), S'\big)\Big) \subseteq \\ \subseteq \mathcal{I}(S')(i) + \mathrm{DIS}\big(\mu, P(i), S'\big) = \mathcal{I}(S')(i)\end{aligned}$$

and thus $\mathcal{I}(S)(i) \subseteq \mathcal{I}(S')(i)$. □

7 A framework construction

In the last couple of sections, the generic theory has been developed to a point where it is comparable to the ring theory diamond lemma *provided* that one can set up the necessary framework. The ideas behind the standard construction have already been presented, but it is convenient to collect everything in a formal statement to facilitate citations in other papers. Furthermore the last couple of sections have demonstrated that one typically wants a bit more than just the basic framework assumptions, so in support of the framework construction there are also some lemmas which give more elementary conditions that suffice for establishing the advanceability, compatibility, and equicontinuity properties.

The first lemma concerns the basic construction of a topology. It serves mainly as a preparation for Lemma 7.6 and may certainly be skipped if one is only interested in a discrete topology. For simplicity, it is stated in single-sorted notation.

Lemma 7.1. *Let $\mathcal{R}$ be an associative unital ring with ring ultranorm $|\cdot|$ which is complete in the topology induced by this norm. Let some nonempty set Y and a function $U\colon Y \longrightarrow \mathbb{R}^+$ be given. Let $\mathcal{M}$ be the free $\mathcal{R}$-module with basis Y. For every $\mu \in Y$, let $f_\mu\colon \mathcal{M} \longrightarrow \mathcal{R}$ be the coefficient-of-μ homomorphism, i.e., $f_\mu(\mu) = 1$ for all $\mu \in Y$ and $f_\mu(\nu) = 0$ for all $\mu, \nu \in Y$ such that $\mu \neq \nu$. Define*

$$\|a\| := \max_{\mu\in Y} \bigl|f_\mu(a)\bigr|\, U(\mu) \tag{7.1}$$

for all $a \in \mathcal{M}$. Then $\|\cdot\|$ is an $\mathcal{R}$-module ultranorm on $\mathcal{M}$. Let

$$\begin{aligned} R &= \bigl\{\, a \mapsto r\cdot a : \mathcal{M} \longrightarrow \mathcal{M} \,\bigm|\, r \in \mathcal{R}, |r| \leqslant 1 \,\bigr\}, \\ B_n &= \bigl\{\, a \in \mathcal{M} \,\bigm|\, \|a\| < 2^{1-n} \,\bigr\} \quad \text{for } n \geqslant 1. \end{aligned}$$

Then $\mathcal{O} = \{B_n\}_{n=1}^\infty$ satisfies Assumption 3 and every f_μ for $\mu \in Y$ is continuous. The extensions of $\{f_\mu\}_{\mu\in Y}$ and $\|\cdot\|$ to $\overline{\mathcal{M}}$ by continuity satisfy (7.1) *and*

$$\bigl\|f_\mu(a)\cdot\mu\bigr\| \leqslant \|a\| \qquad \text{for all } \mu \in Y, \tag{7.2}$$

for all $a \in \overline{\mathcal{M}}$.

Proof. That the right hand side of (7.1) exists for all $b \in \mathcal{M}$ follows from the fact that $\bigl\{\, \mu \in Y \,\bigm|\, f_\mu(b) \neq 0 \,\bigr\}$ is finite for every $b \in \mathcal{M}$. The claim that $\|\cdot\|$ is an $\mathcal{R}$-module ultranorm is shown by verifying the conditions in Definition 2.9. $\|a\|$ is nonnegative because it is a maximum of nonnegative numbers. Since $|\cdot|$ satisfies the strong triangle inequality,

$$\begin{aligned} \|a-b\| = \max_{\mu\in Y} \bigl|f_\mu(a-b)\bigr|\, U(\mu) &\leqslant \max_{\mu\in Y} \max\Bigl\{\bigl|f_\mu(a)\bigr|, \bigl|f_\mu(b)\bigr|\Bigr\} U(\mu) = \\ &= \max\Bigl\{\max_{\mu\in Y} \bigl|f_\mu(a)\bigr|\, U(\mu), \max_{\mu\in Y} \bigl|f_\mu(b)\bigr|\, U(\mu)\Bigr\} = \max\bigl\{\|a\|, \|b\|\bigr\} \end{aligned}$$

for all $a, b \in \mathcal{M}$. $\|a\| = 0$ iff $\big|f_\mu(a)\big| = 0$ for all $\mu \in \mathcal{Y}$, which holds iff $f_\mu(a) = 0$ for all $\mu \in \mathcal{Y}$, which in turn is true iff $a = 0$. For every $r \in \mathcal{R}$ and $a \in \mathcal{M}$,

$$\|r \cdot a\| = \max_{\mu \in Y} \big|f_\mu(r \cdot a)\big| \, U(\mu) = \max_{\mu \in Y} \big|r \cdot f_\mu(a)\big| \, U(\mu) \leqslant$$
$$\leqslant \max_{\mu \in Y} |r| \, \big|f_\mu(a)\big| \, U(\mu) = |r| \, \|a\| \,. \quad (7.3)$$

Hence $\|\cdot\|$ is an $\mathcal{R}$-module ultranorm. As such, it is also a uniformly continuous function $\mathcal{M} \longrightarrow [0, \infty[\subset \mathbb{R}$ since the absolute value of $\|a\| - \|b\|$ by the triangle inequality is bounded from above by $\|a - b\|$. By the completeness of the codomain, it follows that $\|\cdot\|$ extends by continuity to a function $\overline{\mathcal{M}} \longrightarrow [0, \infty[$, and this extended map will also be an $\mathcal{R}$-module ultranorm because left and right hand sides in the axioms for this are all continuous and thus the axioms are preserved under taking limits.

It follows from (7.3) that any set of all $a \in \mathcal{M}$ such that $\|a\| < e$ for some $e \in \mathbb{R}^+$ is an R-module. That $\bigcap_{n=1}^{\infty} B_n = \{0\}$ is because $\|a\| = 0$ implies $a = 0$. Let $e \in \mathbb{R}^+$ and $\mu \in Y$ be given. Since $\big|f_\mu(a)\big| \, U(\mu) \leqslant \|a\|$ for all $a \in \mathcal{M}$, it follows that any $a \in \mathcal{M}$ such that $\|a\| < eU(\mu)$ has $\big|f_\mu(a)\big| < e$, and hence f_μ is continuous. By this continuity and the completeness of $\mathcal{R}$, it extends to a map $\overline{\mathcal{M}} \longrightarrow \mathcal{R}$. Furthermore

$$\big\|f_\mu(a) \cdot \mu\big\| = \max_{\nu \in Y} \Big|f_\nu\big(f_\mu(a) \cdot \mu\big)\Big| \, U(\nu) = \max_{\nu \in Y} \Big|f_\mu(a) \cdot f_\nu(\mu)\Big| \, U(\nu) =$$
$$= \max\Big\{0, \big|f_\mu(a)\big| \, U(\mu)\Big\} = \big|f_\mu(a)\big| \, U(\mu) \leqslant \|a\|$$

which demonstrates (7.2) for $a \in \mathcal{M}$. By continuity of left and right hand sides it continues to hold for arbitrary $a \in \overline{\mathcal{M}}$.

On the matter of (7.1) for $a \in \overline{\mathcal{M}} \backslash \mathcal{M}$, one may first observe that $\|a\| > 0$ and thus there exists some $b \in \mathcal{M}$ such that $\|a - b\| < \|a\|$. It follows from the strong triangle inequality that $\|a\| = \|b\| = \max_{\mu \in Y} \big|f_\mu(b)\big| \, U(\mu)$; let $\mu_0 \in Y$ be an element in which this maximum is attained. Since $\big|f_\mu(c)\big| \, U(\mu) \leqslant \|c\|$ for all $c \in \mathcal{M}$ and $\mu \in Y$, this inequality by continuity holds for all $c \in \overline{\mathcal{M}}$. Hence $\big|f_\mu(a - b)\big| \, U(\mu) \leqslant \|a - b\| < \|a\|$ and thus

$$\big|f_\mu(a)\big| \, U(\mu) = \big|f_\mu(b) + f_\mu(a - b)\big| \, U(\mu) \leqslant$$
$$\leqslant \max\Big\{\big|f_\mu(b)\big| \, U(\mu), \big|f_\mu(a - b)\big| \, U(\mu)\Big\} \leqslant \|a\| \,.$$

On the other hand,

$$\|a\| = \big|f_{\mu_0}(b)\big| \, U(\mu_0) = \big|f_{\mu_0}(a) - f_{\mu_0}(a - b)\big| \, U(\mu_0) \leqslant$$
$$\leqslant \max\Big\{\big|f_{\mu_0}(a)\big| \, U(\mu_0), \big|f_{\mu_0}(a - b)\big| \, U(\mu_0)\Big\}$$

and since $\big|f_{\mu_0}(a-b)\big|\,U(\mu_0) < \|a\|$ it follows that $\big|f_{\mu_0}(a)\big|\,U(\mu_0) = \|a\|$. This has verified not only that the maximum in the right hand side of (7.1) exists, but also that it equals the left hand side. □

The main specialisation made in this construction is that every $\mathcal{M}(i)$ is a free $\mathcal{R}$-module, for some fixed ring $\mathcal{R}$. The topological conditions may seem extensive, but they are all fulfilled in the case considered in Lemma 7.1, and they are of course void in the case of a discrete topology. The classical case is furthermore that $R(i) = \mathcal{R}(i)$ and $\mathcal{Y}(i) = Y(i)$, but as discussed in Section 2, things aren't always that simple.

Construction 7.2. Let $\mathcal{R}$ be a unital associative topologically complete ring. For every $i \in I$, let $Y(i)$ be an arbitrary set and let $\mathcal{M}(i)$ be the free $\mathcal{R}$-module with basis $Y(i)$. For every $\mu \in Y(i)$, denote by f_μ the coefficient-of-μ homomorphism $\mathcal{M}(i) \longrightarrow \mathcal{R}$. Also let $\mathcal{R}(i)$ be the ring of $\mathcal{R}$-actions on $\mathcal{M}(i)$, let $R(i)$ be a subring of $\mathcal{R}(i)$, and let $R^\perp(i) \subseteq \mathcal{R}(i)$ be a set such that $\sum_{r\in R^\perp(i)} R(i)\circ r = \mathcal{R}(i)$ and $\mathrm{id} \in R^\perp(i)$. Let $\mathcal{Y}(i) = \big\{\, r(\mu) \,\big|\, r \in R^\perp(i), \mu \in Y(i) \,\big\}$; this ensures Assumption 2 is fulfilled. For every $i \in I$, let $\mathcal{O}(i) = \big\{B_n(i)\big\}$ be a family of $R(i)$-modules satisfying Assumption 3 and in addition being such that the $\mathcal{R}$-module multiplication $\mathcal{R} \times \overline{\mathcal{M}}(i) \longrightarrow \overline{\mathcal{M}}(i) : (r, b) \mapsto r \cdot b$ and all maps $\{f_\mu\}_{\mu\in Y(i)}$ are continuous.

Let $V = \bigcup_{i,j\in I} V(i,j)$ be such that every $v \in V(i,j)$ is a continuous $\mathcal{R}$-linear map $\overline{\mathcal{M}}(j) \longrightarrow \overline{\mathcal{M}}(i)$. Let $S = \bigcup_{i\in I} S(i)$ be arbitrary such that $S(i) \subseteq Y(i) \times \overline{\mathcal{M}}(i)$. For all $i, j \in I$, let

$$W(i,j) = \Big\{\, \big(v, (\mu, a)\big) \in V(i,j) \times S(j) \,\Big|\, v(\mu) \in Y(i) \,\Big\}$$

and define $T_1(S)(i) = \bigcup_{j\in I}\{t_{v,s}\}_{(v,s)\in W(i,j)}$, where

$$t_{v,(\mu,a)}(b) = b - f_{v(\mu)}(b)\cdot v(\mu - a) \qquad \text{for all } b \in \overline{\mathcal{M}}(i). \tag{7.4}$$

This $T_1(S)(i)$ satisfies Assumption 4 and every $r \in \mathcal{R}(i)$ is absolutely advanceable with respect to $T_1(S)(i)$, for all $i \in I$. Furthermore every ambiguity $\big(t_1, r(\mu), t_2\big)$ of $T_1(S)(i)$ where $\mu \in Y(i)$ and $r \in R^\perp(i) \setminus \{\mathrm{id}\}$ is an absolute shadow of the ambiguity (t_1, μ, t_2).

Proof. Let $i \in I$ be given. That any $t \in T_1(S)(i)$ is continuous and $\mathcal{R}$-linear follows from (7.4) since this formula is a composition of maps with these properties, and thus t is a continuous homomorphism satisfying $t \circ r = r \circ t$ for all $r \in \mathcal{R}(i)$. Not only does this satisfy Assumption 4, but it also means every $r \in \mathcal{R}(i)$ is absolutely advanceable. □

Lemma 7.3. *Let everything be as in the construction. Assume in addition that V is closed under composition and satisfies $v\big(Y(j)\big) \subseteq Y(i) \cup \{0\}$ for all $v \in V(i,j)$ and $i, j \in I$. If S is such that $v(\mu) = 0$ implies $v(a) = 0$ for $v \in V(i,j)$, $(\mu, a) \in S(j)$, and $i, j \in I$ then for all $i, j \in I$ every element of $V(i,j)$ is advanceable with respect to $T_1(S)(j)$ and $T_1(S)(i)$.*

Proof. Let $i, j \in I$, $w \in V(i,j)$ such that $w\big(Y(j)\big) \subseteq Y(i) \cup \{0\}$, $t' \in T_1(S)(j)$, and $b \in R^*\mathcal{Y}(j)$ be given. By definition of $T_1(S)(j)$ there exists some $k \in I$ and $\big(v, (\mu, a)\big) \in W(j,k)$ such that $t' = t_{v,(\mu,a)}$. Furthermore there exist $\nu \in Y(j)$ and $r \in \mathcal{R}$ such that $b = r \cdot \nu$. There are three cases for $v\big(t'(b)\big)$.

1. If $v(\mu) \neq \nu$ then $f_{v(\mu)}(b) = 0$ and hence $t'(b) = b$, in which case $t\big(w(b)\big) = w\big(t'(b)\big)$ for $t = \mathrm{id}$.

2. If $v(\mu) = \nu$ and $w(\nu) \in Y(i)$ then $\big(w \circ v, (\mu, a)\big) \in W(i,k)$ and hence one can consider $t = t_{w \circ v,(\mu,a)}$, which is the most interesting case. $f_{v(\mu)}(b) = r$ and hence $t'(b) = r \cdot v(a)$, so that $w\big(t'(b)\big) = r \cdot w\big(v(a)\big) = r \cdot t\big((w \circ v)(\mu)\big) = t\big(r \cdot w(\nu)\big) = t\big(w(b)\big)$, as claimed.

3. If $v(\mu) = \nu$ and $w(\nu) = 0$ then $(w \circ v)(\mu) = 0$ and hence $(w \circ v)(a) = 0$, which means $w\big(t'(b)\big) = w\big(r \cdot v(a)\big) = r \cdot (w \circ v)(a) = r \cdot (w \circ v)(\mu) = w\big(r \cdot v(\mu)\big) = w(b)$, and thus $t\big(w(b)\big) = w\big(t'(b)\big)$ for $t = \mathrm{id}$.

Either way, there exists some $t \in T(S)(i)$ such that $t\big(w(b)\big) = w\big(t'(b)\big)$. □

Lemma 7.4. *Let everything be as in the construction. Let a partial order $P(i)$ on $\mathcal{Y}(i)$ be given for every $i \in I$. Assume every $v \in V(i,j)$ correlates $P(j)$ to $P(i)$ and every $r \in R^\perp(i)$ correlates $P(i)$ to itself. If $a \in \mathrm{DSM}\big(\mu, P(j)\big)$ for all $(\mu, a) \in S(j)$ and $j \in I$, then $T(S)(i)$ is compatible with $P(i)$ for all $i \in I$.*

Proof. Let $i, j \in I$, $\big(v, (\mu, a)\big) \in W(i,j)$, and $\nu \in \mathcal{Y}(i)$ be given. It must be shown that $t_{v,(\mu,a)}(\nu) \in \{\nu\} \cup \mathrm{DSM}\big(\nu, P(i)\big)$. If $f_{v(\mu)}(\nu) = 0$ then $t_{v,(\mu,a)}(\nu) = \nu$ and all is well. Otherwise $\nu = r \cdot v(\mu)$ for $r = f_{v(\mu)}(\nu)$ and $t_{v,(\mu,a)}(\nu) = r \cdot v(a)$. Since $a \in \mathrm{DSM}\big(\mu, P(j)\big)$ it follows that $v(a) \in v\big(\mathrm{DSM}\big(\mu, P(j)\big)\big) \subseteq \mathrm{DSM}\big(v(\mu), P(i)\big)$ and hence $r \cdot v(a) \in r \cdot \mathrm{DSM}\big(v(\mu), P(i)\big) \subseteq \mathrm{DSM}\big(\nu, P(i)\big)$. Therefore all $t \in T_1(S)(i)$ are compatible with $P(i)$. By Lemma 5.4, this extends to the whole of $T(S)(i)$. □

Lemma 7.5. *Let everything be as in the construction. Let $i \in I$ be given and let $\mathcal{N} \supseteq \widehat{\mathcal{O}}(i)$ be a family of topologically open subgroups of $\overline{\mathcal{M}}(i)$ such that $f_\nu(a) \cdot \nu \in \varepsilon$ for all $a \in \varepsilon$, $\varepsilon \in \mathcal{N}$, and $\nu \in Y(i)$. If all $j \in I$, $\big(v, (\mu, a)\big) \in W(i,j)$, $r \in \mathcal{R}$, and $\varepsilon \in \mathcal{N}$ such that $r \cdot v(\mu) \in \varepsilon$ also satisfy $r \cdot v(a) \in \varepsilon$ then:*

1. *$t(\varepsilon) \subseteq \varepsilon$ for any $\varepsilon \in \mathcal{N}$ and $t \in T(S)(i)$.*

2. *$T(S)(i)$ is equicontinuous.*

Proof. The second claim is an immediate consequence of the first ($\delta = \varepsilon$ works for all reductions). If the first claim holds for two particular reductions, then it also holds for their composition; hence it is sufficient to

verify it for simple reductions. Let $t \in T_1(S)(i)$ be given. By definition there is some $j \in I$ and $\big(v, (\mu, a)\big) \in W(i,j)$ such that $t = t_{v,(\mu,a)}$, i.e., $t(b) = b - f_{v(\mu)}(b) \cdot v(\mu - a)$ for all $b \in \overline{\mathcal{M}}(i)$. Let $\varepsilon \in \mathcal{N}$ and $b \in \varepsilon$ be arbitrary. $f_{v(\mu)}(b) \cdot v(\mu) \in \varepsilon$ by the condition on $\mathcal{N}$ and hence $f_{v(\mu)}(b) \cdot v(a) \in \varepsilon$ by assumption. It follows that $t(b) = b - f_{v(\mu)}(b) \cdot v(\mu) + f_{v(\mu)}(b) \cdot v(a) \in \varepsilon - \varepsilon + \varepsilon = \varepsilon$. □

In [7, Lemma 3.25], a different proof of equicontinuity can be found which is feasible also in cases where the first conclusion of the above lemma does not hold; the idea is to consider $T(S)(i)$ that are compatible with some $P(i)$ and require the latter to satisfy a 'squeeze property' (as in the *Squeeze Theorem* of elementary analysis): for every $\varepsilon \in \widehat{\mathcal{O}}(i)$ there must exist some $\delta \in \widehat{\mathcal{O}}(i)$ such that if $\mu \in \mathcal{Y}(i) \cap \varepsilon$ then every $\nu < \mu$ in $P(i)$ must satisfy $\nu \in \delta$. However, I currently don't have any example of a situation where this additional generality is needed. That proof is also easily disturbed by the existence of "small" scalars, since it might happen that the $r(\mu) \in \varepsilon \cap R^* \mathcal{Y}(i)$ some $t \in T(S)(i)$ acts upon does not satisfy $\mu \in \varepsilon$; it is typically necessary to have some condition ensuring that elements of $R(i)$ act somewhat uniformly on $\overline{\mathcal{M}}(i)$, and even then things can get hairy.

The final lemma is instead a special case of Lemma 7.5 which separates the norm conditions on S, V, and $\mathcal{R}$.

Lemma 7.6. *Let everything be as in the construction. Let a function $U_i \colon Y(i) \longrightarrow \mathbb{R}^+$ be given for every $i \in I$. Assume the topology in $\mathcal{R}$ is given by a ring ultranorm $|\cdot|$. Also assume for all $i \in I$ that $\mathcal{O}(i)$ is constructed from $|\cdot|$ and $U(i)$ as in Lemma 7.1, and let $\|\cdot\|_i$ be the $\mathcal{R}$-module ultranorm on $\overline{\mathcal{M}}(i)$. Assume that there for every $v \in V$ exists a constant $C_v \in \mathbb{R}^+$ such that if $v \in V(i,j)$ and $\mu \in Y(j)$ then $\big\|v(\mu)\big\|_i \leqslant C_v U_j(\mu)$ and if in addition $v(\mu) \in Y(i)$ then $U_i\big(v(\mu)\big) = C_v U_i(\mu)$.*

If $\|a\|_j \leqslant U_j(\mu)$ for all $(\mu, a) \in S(j)$ and $j \in I$, then $\big\|t(b)\big\|_i \leqslant \|b\|_i$ for all $t \in T(S)(i)$, $b \in \overline{\mathcal{M}}(i)$, and $i \in I$, and moreover $T(S)(i)$ is equicontinuous for every $i \in I$.

Proof. This is a special case of Lemma 7.5. In order for $t(\varepsilon) \subseteq \varepsilon$ for any $\varepsilon \in \mathcal{N}$ and $t \in T(S)(i)$ to imply $\big\|t(b)\big\|_i \leqslant \|b\|_i$ for all $t \in T(S)(i)$ and $b \in \overline{\mathcal{M}}(i)$, it is necessary to take

$$\mathcal{N} = \left\{ \left\{ b \in \overline{\mathcal{M}}(i) \;\middle|\; \|b\|_i < e \right\} \;\middle|\; e \in \mathbb{R}^+ \right\},$$

but apart from that the proof is purely a matter of demonstrating that the conditions in Lemma 7.5 are met. That $f_\nu(a) \cdot \nu \in \varepsilon$ for all $a \in \varepsilon$, $\varepsilon \in \mathcal{N}$, and $\nu \in Y(i)$ follows from (7.2).

For the main condition, let $i, j \in I$, $\big(v, (\mu, a)\big) \in W(i,j)$, $r \in \mathcal{R}$, and $e \in \mathbb{R}^+$ such that $\big\|r \cdot v(\mu)\big\|_i < e$ be given. By (7.1), $\big\|r \cdot v(\mu)\big\|_i = |r|\, U_i\big(v(\mu)\big) =$

$|r|\, C_v U_j(\mu)$. For every $\nu \in Y(j)$,

$$\begin{aligned}\Big\|r \cdot v\big(f_\nu(a)\cdot\nu\big)\Big\|_i &\leqslant |r|\,\big\|f_\nu(a)\cdot v(\nu)\big\|_i \leqslant |r|\,\big|f_\nu(a)\big|\,\big\|v(\nu)\big\|_i \leqslant \\ &\leqslant |r|\,\big|f_\nu(a)\big|\,C_v U_j(\nu) \leqslant |r|\,C_v\,\|a\|_j \leqslant |r|\,C_v U_j(\mu) = \big\|r\cdot v(\mu)\big\|_i < e.\end{aligned}$$

Since $\|\cdot\|_i$ is an ultranorm and $r \cdot v(a)$ is in the topologically closed group generated by $\big\{r\cdot v\big(f_\nu(a)\cdot\nu\big)\big\}_{\nu\in Y(j)}$, it now follows that $\big\|r\cdot v(a)\big\|_i < e$. □

8 Gröbner bases

The following treatment of Gröbner bases is primarily aimed at demonstrating how some known results in this area can be derived from the diamond lemma, hence it does not seek to give a definition of Gröbner basis that applies in all situations covered by the $\big(\mathcal{M}, R, \mathcal{Y}, \mathcal{O}, T_1(S)\big)$ formalism. The restrictions that will be made are:

1. There will only be one sort.
2. The topology will be discrete.
3. $\mathcal{M}$ will be a free left $\mathcal{R}$-module, where $\mathcal{R}$ is a unital ring, and R will be the set of maps that multiply by an element of $\mathcal{R}$.
4. $\mathcal{Y}$ will be a basis of $\mathcal{M}$.

One restriction that will *not* be made is that of only considering total orders, as that is needed more to ensure existence of Gröbner bases than to define or use them. Some algebraic structures require compatible partial orders to be non-total, so a restriction to total orders really sacrifices some generality.

8.1 Generic theory

On a practical level, the property that something is a Gröbner basis is equivalent to the four claims in Theorem 5.11, which means several equivalent characterisations of this concept could be made. The standard definition is however the fifth claim that '*the leading monomial of an element of the ideal must be a multiple of the leading monomial of some element of the basis*', which accordingly appears as claim (d) of Theorem 8.4. One reason this characterisation has become so popular is no doubt that it is amenable to an informal presentation — everybody knows what the leading monomial is, don't they? — although once one starts to do anything with the concept (such as reducing modulo a tentative Gröbner basis), most technical details of a reduction-based approach quickly suggest themselves. Moreover, even the issue of what it means to be the leading monomial is not without technical complications when considered in the present generality.

Definition 8.1. Let $\{f_\mu \colon \mathcal{M} \longrightarrow \mathcal{R}\}_{\mu\in\mathcal{Y}}$ be the family of coefficient-of-μ homomorphisms associated with the basis $\mathcal{Y}$ for $\mathcal{M}$. The **support** $\operatorname{supp}(a)$ of an $a \in \mathcal{M}$ is the set of $\mu \in \mathcal{Y}$ for which $f_\mu(a) \neq 0$.

Let P be a binary relation on $\mathcal{Y}$. A **P-leading monomial** of some $a \in \mathcal{M}$ is a P-maximal element of $\operatorname{supp}(a)$, i.e., a $\mu \in \operatorname{supp}(a)$ such that no $\nu \in \operatorname{supp}(a)$ satisfies $\mu < \nu$ in P. Denote by $\operatorname{LM}_P(a)$ the set of P-leading monomials of a. If $\operatorname{LM}_P(a)$ has exactly one element, then denote that by $\operatorname{lm}_P(a)$.

The main reason for restricting this treatment to $\mathcal{Y}$ being a basis of $\mathcal{M}$ and the topology being discrete is that this ensures $\operatorname{supp}(a)$ — informally "the set of monomials occurring in a" — is well-defined. Linear dependencies in $\mathcal{Y}$ would obviously remove the foundation for this concept, and topology can (but doesn't have to) produce similar problems: on can choose a topology such that there are $Y_1, Y_2 \subset \mathcal{Y}$ for which $\operatorname{Span}(Y_1) \cap \operatorname{Span}(Y_2) = \{0\}$ but $\operatorname{Cspan}(Y_1) \cap \operatorname{Cspan}(Y_2) \neq \{0\}$. Even after ensuring that $\operatorname{supp}(a)$ is well-defined for all $a \in \overline{\mathcal{M}}$, a topology can cause the definition of $\operatorname{LM}_P(a)$ to fail, if some $\operatorname{supp}(a)$ is infinite and contains an infinite P-ascending chain. One approach for defining Gröbner bases without relying on the support concept could be to replace the concrete construction of $\operatorname{LM}_P(g)$ by an abstract map L that assigns a set of leading monomials to each element of $\overline{\mathcal{M}}$. The effect would probably be similar to the formalism in [11], even though that technically goes in the other direction: the "L" map has a canonical construction but the monomials are abstracted away.

Definition 8.2. Let V be a set of $\mathcal{R}$-module homomorphisms $\mathcal{M} \longrightarrow \mathcal{M}$. Let P be a binary relation on $\mathcal{Y}$. Let $N \subseteq \mathcal{M}$ be a V-ideal. A subset G of $\mathcal{M}$ is said to be a **P-monic V-Gröbner-basis** of N if it is P-monic, $v(g) \in N$ for all $v \in V$ and $g \in G$, and there for every $a \in N$ and $\mu \in \operatorname{LM}_P(a)$ exists some $g \in G$ and $v \in V$ such that $\mu = v\big(\operatorname{lm}_P(g)\big)$.

Making P-monicity a precondition for Gröbner bases serves two purposes: it ensures there is a unique P-leading monomial and it ensures reductions compatible with P can be manufactured from basis elements. While these are important ingredients in Lemma 8.3, the P-monicity condition also works against a very elementary result in traditional Gröbner basis theory, namely that every ideal should have a Gröbner basis. Without the P-monicity it would be possible to simply make the observation that the ideal itself is a Gröbner basis for it — admittedly a ridiculously large basis (probably infinite in most cases where finite bases exist), but nonetheless a basis — which formally justifies assuming every ideal one needs to work with is generated by a Gröbner basis. If a Gröbner basis is to be P-monic however, one has to be careful about what elements can be included, but as long as P is a total order and $\mathcal{R}$ is a field there is always a P-monic counterpart of every nonzero element of $\mathcal{M}$.

It should also be observed that the definition of Gröbner basis does not explicitly require G to be a V-ideal basis for N, and in fact it depends on P whether this follows. A trivial counterexample is to consider $N = \mathcal{M} = \mathcal{R}[\mathsf{x}]$ and $G = \{1 + \mathsf{x}\}$ where $\mathcal{R}$ is a field, $\mathcal{Y} = \{\mathsf{x}^n\}_{n\in\mathbb{N}}$, $V = \{b \mapsto \mathsf{x}^n b\}_{n\in\mathbb{N}}$, and $\mathsf{x}^m \leqslant \mathsf{x}^n$ in P iff $m \geqslant n$; since $\mathrm{lm}_P(1 + \mathsf{x}) = 1$ it is easy to see that G is a V-Gröbner-basis for N, but $1 \notin \mathrm{Span}\big(\{(1 + \mathsf{x})\mathsf{x}^n\}_{n\in\mathbb{N}}\big)$ and so G isn't a V-ideal basis for N. The catch in this example is that P isn't well-founded; the condition defining Gröbner bases lends itself to the step in an induction for proving the ideal basis property, but it cannot also provide the base for that induction. The next lemma gives sufficient conditions on V and P for Gröbner bases to be ideal bases.

Lemma 8.3. *Let V be a monoid of $\mathcal{R}$-module homomorphisms $\mathcal{M} \longrightarrow \mathcal{M}$. Let P be a well-founded partial order on $\mathcal{Y}$ which is correlated to itself by every $v \in V$. If $G \subseteq \mathcal{M}$ is a P-monic V-Gröbner-basis for a V-ideal $N \subseteq \mathcal{M}$ then G is a V-ideal basis for N. If furthermore $T_1(S)$ is as in Construction 7.2 for $S = \big\{\big(\mathrm{lm}_P(g), \mathrm{lm}_P(g) - g\big)\big\}_{g\in G}$ then $\mathcal{I}(S) = N$ and for every $a \in N$ there exists some $t \in T(S)$ such that $t(a) = 0$.*

Proof. The main claim is that about existence of reductions which map elements of N to 0. What the Gröbner property implies is that there for every nonzero $a \in N$ and $\mu \in \mathrm{LM}_P(a)$ exists some $t \in T_1(S)$ which acts nontrivially on μ, namely $t = t_{\mu\mapsto\mu-v(g)}$ where $g \in G$ and $v \in V$ are such that $\mu = v\big(\mathrm{lm}_P(g)\big)$, since this is $t_{v,(\nu,b)}$ where $\nu = \mathrm{lm}_P(g)$ and $b = \nu - g$. All these simple reductions are compatible with P, since $b \in \mathrm{DSM}(\nu, P)$ because g is P-monic and $v(b) \in \mathrm{DSM}(\mu, P) = \mathrm{DSM}\big(v(\nu), P\big)$ by assumption.

Let $a_0 \in N$ be given. Construct from any nonzero $a_n \in N$ the element $a_{n+1} = u_n(a_n)$ by picking as u_n some composition $u_{n,m_n} \circ \cdots \circ u_{n,1}$ of simple reductions such that $u_{n,k} \in T_1(S)$ acts nontrivially on $\mu_{n,k} \in \mathrm{LM}_P(a_n)$, where $\{\mu_{n,1}, \ldots, \mu_{n,m_n}\} = \mathrm{LM}_P(a_n)$. The claim follows once it has been shown that $a_n = 0$ for some n (which means $u_l = \mathrm{id}$ for all $l > n$), and the way to establish this is to consider the sets $\mathrm{LM}_P(a_n)$.

Let $Z = \bigcup_{n=0}^{\infty} \mathrm{LM}_P(a_n)$. First observe that $u_n(\mu_{n,k}) \in \mathrm{DSM}(\mu_{n,k}, P)$ for any n and k by Lemma 5.4. Since any P-leading monomial of a_n is some $\mu_{n,k}$, and since $\mathrm{Span}\big(\mathrm{supp}(a_n) \setminus \mathrm{LM}_P(a_n)\big) \subseteq \sum_{k=1}^{m_n} \mathrm{DSM}(\mu_{n,k}, P)$, it follows that

$$\mathrm{LM}_P(a_{n+1}) \subseteq \mathrm{supp}(a_{n+1}) \subseteq \big\{\, \nu \in \mathcal{Y} \,\big|\, \nu < \mu \text{ in } P \text{ for some } \mu \in \mathrm{LM}_P(a_n) \,\big\}$$

for $n = 0, 1, \ldots$. Construct the directed acyclic graph D which has Z as vertex set and has an edge from μ to ν iff $\mu > \nu$ in P and there exists some $n \in \mathbb{N}$ such that $\mu \in \mathrm{LM}_P(a_n)$ and $\nu \in \mathrm{LM}_P(a_{n+1})$. Since any path in this graph is a P-descending chain, it is finite. Since any $\mathrm{LM}_P(a_n)$ is finite, the graph has finite branching. Finally the only roots in D are the elements of $\mathrm{LM}_P(a_0)$. Hence König's lemma (an infinite tree with finite branching has

an infinite path) applies, and it follows that Z is finite. In particular, there exists some n for which $\mathrm{LM}_P(a_n) = \varnothing$ and thus $a_n = 0$, as claimed.

That $N \subseteq \mathcal{I}(S)$ is now immediate from the definition of the latter. Conversely it may be observed that if $\mu - t(\mu) \neq 0$ for some $t \in T_1(S)$ and $\mu \in \mathcal{Y}$ then there exist $g \in G$ and $v \in V$ such that $\mu - t(\mu) = v(g)$, since t is of the form $t_{\mu \mapsto \mu - v(g)}$. Thus

$$\mathcal{I}(S) = \mathrm{Cspan}\Big(\big\{\, v(\mu - a) \,\big|\, v \in V, (\mu, a) \in S, v(\mu) \in \mathcal{Y} \,\big\}\Big) \subseteq$$
$$\subseteq \mathrm{Cspan}\Big(\big\{\, v(g) \,\big|\, v \in V, g \in G \,\big\}\Big) \subseteq N$$

and this also shows that G is a V-ideal basis of N. □

With this result about the existence of reductions which map ideal elements to 0, it becomes easy to link the Gröbner basis concept to those of Theorem 5.11. Claim (a‴) below is included because it is literally the claim that "all S-polynomials reduce to 0" which is practically used to verify that something is a Gröbner basis.

Theorem 8.4. *Let P be a well-founded partial order on $\mathcal{Y}$. Let V be a monoid of $\mathcal{R}$-module homomorphisms $\mathcal{M} \longrightarrow \mathcal{M}$ that map $\mathcal{Y}$ into $\mathcal{Y}$ and are strictly monotone with respect to P.*

Let $S \subseteq \mathcal{Y} \times \mathcal{M}$ be such that $a \in \mathrm{DSM}(\mu, P)$ for any $(\mu, a) \in S$. Let $T_1(S)$ be as in Construction 7.2. Then the following conditions are equivalent:

(a) *Every ambiguity of $T_1(S)$ is resolvable.*

(a′) *Every ambiguity of $T_1(S)$ is resolvable relative to P.*

(a″) *Every V-critical ambiguity of $T_1(S)$ is resolvable relative to P.*

(a‴) *For every V-critical ambiguity (t_1, μ, t_2) of $T_1(S)$ there exists some $t \in T(S)$ such that $t\big(t_1(\mu) - t_2(\mu)\big) = 0$.*

(b) *Every element of $\mathcal{M}$ is persistently and uniquely reducible.*

(c) *Every element of $\mathcal{M}$ has a unique normal form.*

(d) *The set $\{\mu - a\}_{(\mu,a)\in S}$ is a P-monic V-Gröbner-basis of $\mathcal{I}(S)$.*

Proof. First observe that strict monotonicity of V implies correlation by Lemma 6.5, and hence $T(S)$ is compatible with P by Lemma 7.4. All elements of V are advanceable with respect to $T_1(S)$ by Lemma 7.3, and thus $\mathcal{I}(S)$ is a V-ideal by Lemma 3.7.

Claims (a), (a′), (b), and (c) are equivalent by Theorem 5.11. Claims (a′) and (a″) are equivalent by Theorem 6.9. (a‴) implies that every V-critical ambiguity is resolvable, and hence (a″) by Lemma 5.10. Conversely

(b) implies that $\mu \in \mathrm{Red}(S)$ for every V-critical ambiguity (t_1, μ, t_2) of $T_1(S)$ and hence $t^S\big(t_1(\mu)\big) = t^S\big(t_2(\mu)\big)$, from which follows $t^S\big(t_1(\mu) - t_2(\mu)\big) = 0$ and thus (a$'''$).

Assume (d). By Lemmas 5.5 and 4.2, $\mathcal{M} = \mathrm{Irr}(S) + \mathcal{I}(S)$. Furthermore every $a \in \mathcal{I}(S) \cap \mathrm{Irr}(S)$ satisfies $t(a) = a$ for all $t \in T(S)$, but by Lemma 8.3 there is some $t \in T(S)$ such that $t(a) = 0$. Hence $a = 0$, which has established $\mathcal{M} = \mathrm{Irr}(S) \oplus \mathcal{I}(S)$. It follows that claim (d) implies claim (c).

Finally assume $\mathrm{Red}(S) = \mathcal{M}$ and let $b \in \mathcal{I}(S)$ be arbitrary. Since $t^S(b) = 0$ there exists some $t \in T(S)$ such that $t(b) = 0$. Let $\lambda \in \mathrm{LM}_P(b)$ be arbitrary. Since $\lambda \notin \mathrm{LM}_P(0)$, there exists some decomposition $t = t_3 \circ t_2 \circ t_1$ where $t_3, t_1 \in T(S)$ and $t_2 \in T_1(S)$ are such that $\lambda \in \mathrm{LM}_P\big(t_1(b)\big)$ but $\lambda \notin \mathrm{LM}_P\big((t_2 \circ t_1)(b)\big)$. Since there is no $\nu \in \mathrm{supp}\big(t_1(b)\big)$ such that $\nu > \lambda$ in P, it must be the case that t_2 acts nontrivially on λ. Due to the way $T_1(S)$ was constructed, this means there is some $v \in V$ and $(\mu, a) \in S$ such that $v(\mu) = \lambda$. Since furthermore $\mu = \mathrm{lm}_P(\mu - a)$, the condition with respect to b and λ for $\{\mu - a\}_{(\mu,a)\in S}$ to be a V-Gröbner-basis is fulfilled. Hence claim (b) implies claim (d). □

A classical case not handled by this theorem is that P is a total order but elements of $\mathcal{R}$ sometimes aren't invertible. This is where monicity becomes a restriction, since there in for example the case that $\mathcal{R}$ is an euclidian domain exists an established theory — which in addition to gaussian elimination and polynomial division also generalises integer division (with remainder) — for Gröbner bases where leading terms may have noninvertible coefficients. Reducing a term $a\mu\nu$ by a basis element g whose leading term is $b\mu$ then consists of subtracting $qg\nu$ from $a\mu\nu$, where q is the quotient of a by b, and may thus fail to completely eliminate the $\mu\nu$ term. What makes this incompatible with the diamond lemma framework is however that the quotient q, and hence the reduction as a whole, is not given by a homomorphism; $(a_1 \operatorname{div} b) + (a_2 \operatorname{div} b)$ is not always equal to $(a_1 + a_2) \operatorname{div} b$, even through the error may be predictable. The standard bases formalism of [11, 13] has facilities[1] that can handle this, and it's quite possible that the same trick could be applied also in a modification of the diamond lemma formalism, but for the moment I don't see a pressing need for this. It is more interesting to examine some alternative approaches for coping with noninvertible coefficients within the present framework, even though they are perhaps not as general.

[1] In particular the duplication of addition operations: one which is "with carry" (coming from the filtered structure) and one which is "without carry" (coming from the associated graded structure). For a suitable choice of filtered structure the latter addition has $(a_1 \operatorname{div} b) + (a_2 \operatorname{div} b) = (a_1 + a_2) \operatorname{div} b$, and since reductions are required to be homomorphisms with respect to this "without carry" addition, it is then allowed to use integer division when constructing reductions.

If $\mathcal{R}$ can be regarded as an algebra over some smaller subring (maybe even subfield) $\mathcal{K}$, then a practical solution can be to change the boundary between R and $\mathcal{Y}$, letting the former encode just $\mathcal{K}$ and extending $\mathcal{Y}$ accordingly. Corollary 8.6 below can be viewed as using this approach to Gröbner bases in $\mathcal{R}\langle X\rangle$ where $\mathcal{R}$ itself is a commutative polynomial algebra $\mathcal{K}[X_1]$.

The other extreme is that $\mathcal{R} = \mathbb{Z}$, in which case there is no additional freedom that can be gained from shrinking R since the group structure alone determines what multiples of monomials are mapped to. Consider the case that one wishes to make a model for $\mathbb{Z}[x]/\langle 2x\rangle$. It is easy to jump to the conclusion that the diamond lemma framework cannot handle this, on the grounds that $\mathcal{M} = \mathbb{Z}[x]$ is a free $\mathbb{Z}$-module and hence any $\mathrm{Irr}(S)$ must be free too whereas the target $\mathcal{M}/\mathcal{I}(S) = \mathbb{Z}[x]/\langle 2x\rangle$ clearly is not. It is true that Theorem 8.4 is so restricted, but there is nothing in the generic theory which requires one to pick $\mathbb{Z}[x]$ as $\mathcal{M}$ (even though that would be the trivial choice). An interesting alternative in this case would be $\mathcal{M} = \mathbb{Z}[x] \oplus \mathbb{Z}_2[x_2]$, since one for $T_1(S) = \{t_{x^n \mapsto x_2^n}\}_{n=1}^{\infty} \cup \{t_{x_2^0 \mapsto 0}\}$ indeed gets $\mathbb{Z}[x]/\langle 2x\rangle \cong \mathcal{M}/\mathcal{I}(S) \cong \mathrm{Irr}(S)$ as $\mathbb{Z}$-modules.

This approach of introducing "modular coefficients" in parallel with the original coefficients will however not work for the formally similar case of $\mathbb{Z}[x]/\langle 2x - 1\rangle$. Whereas a map that for all $n \in \mathbb{Z}$ takes $2nx$ to n and $(2n+1)x$ to $n + x_2$ makes sense as a map (and indeed is what one would arrive at in the standard bases formalism), it cannot serve as a reduction in the diamond lemma formalism because it is not a group homomorphism; $x_2 + x_2 = 0 \neq 1$. What *will* work is instead to pick $\mathcal{M} = \mathbb{Z}[\frac{1}{2}][x]$, where $\mathbb{Z}[\frac{1}{2}]$ should be regarded as the subring of $\mathbb{Q}$ generated by $\frac{1}{2}$. The main justification for introducing such a multiplicative inverse of 2 would be the defining identity $2x \equiv 1$ itself — whose interpretation must be that x is precisely such an inverse — and once $\frac{1}{2}$ is available the rest is trivial.

The common idea generalising both cases appears to be that if one wants to make a reduction mapping $r\mu$ to a and r is neither invertible nor a zero divisor, then one should extend $\mathcal{M}$ with a new element μ' such that $r\mu' = a$ and have the reduction map μ to μ'. (The 'new' is important here, because $\mathbb{Z}[x,y]/\langle 2x - 2y\rangle \cong \mathbb{Z}[y] \oplus \xi\mathbb{Z}_2[\xi, y]$ where $\xi = y - x'$; that $2x' = 2y$ but $x' \neq y$ since x' is new is what creates the characteristic 2 part.) Whether this method can be turned into an algorithm (as has been done for Gröbner bases over e.g. euclidian domains) is at the time of writing unclear — automating this kind of modifications to the base group $\mathcal{M}$ seems highly nontrivial — but it should illustrate the usefulness of not having Theorem 5.11 restricted to the case of $\mathcal{M}$ being a free module, even though that assumption simplifies the results in this section quite considerably.

8.2 Commutative, associative, and nonassociative algebras

Many forms of the fundamental theorem of Gröbner bases turn out to be special cases, with particular choices of $\mathcal{M}$ and V, of Theorem 8.4 and therefore follow from it as mere corollaries. The most classical is that for commutative polynomials over a field.

Corollary 8.5 (Buchberger)**.** *Let $\mathcal{R}$ be a field, let X be a set, let $\mathcal{M} = \mathcal{R}[X]$, and let $\mathcal{Y}$ be the set of monomials (power products) in $\mathcal{M}$. Let $V = \{b \mapsto \mu b\}_{\mu\in\mathcal{Y}}$ (a set of maps $\mathcal{M} \longrightarrow \mathcal{M}$). Let P be a well-founded semigroup total order on $\mathcal{Y}$. Define a map $Z\colon \big(\mathcal{Y} \times \mathcal{M}\big)^2 \longrightarrow \mathcal{M}$ by*

$$Z\big((\mu_1, a_1), (\mu_2, a_2)\big) = \frac{\operatorname{lcm}(\mu_1, \mu_2)}{\mu_1} a_1 - \frac{\operatorname{lcm}(\mu_1, \mu_2)}{\mu_2} a_2 \tag{8.1}$$

where $\operatorname{lcm}(\mu_1, \mu_2)$ denotes the least common multiple of μ_1 and μ_2.

Let $S \subseteq \mathcal{Y} \times \mathcal{M}$ be such that $a \in \operatorname{DSM}(\mu, P)$ for every $(\mu, a) \in S$ and let $T_1(S)$ be as in Construction 7.2. Then the following are equivalent:

1. *$\{\mu - a\}_{(\mu,a)\in S}$ is a P-monic V-Gröbner basis of $\mathcal{I}(S)$.*
2. *For every pair $(s_1, s_2) \in S^2$ there exists some $t \in T(S)$ such that $t\big(Z(s_1, s_2)\big) = 0$.*
3. *For every $s_1 = (\mu_1, a_1) \in S$ and $s_2 = (\mu_2, a_2) \in S$ such that μ_1 and μ_2 are not coprime there exists some $t \in T(S)$ such that $t\big(Z(s_1, s_2)\big) = 0$.*

Proof. This is mostly the equivalence of (d) and (a‴) from Theorem 8.4, but there are minor variations so it doesn't hurt to make the chain of implications explicit.

The first claim is exactly (d), so this is equivalent to $\operatorname{Red}(S) = \mathcal{M}$. Let $s_1 = (\mu_1, a_1) \in S$ and $s_2 = (\mu_2, a_2) \in S$ be arbitrary. Let $\nu_1 = \operatorname{lcm}(\mu_1, \mu_2)/\mu_1$ and $\nu_2 = \operatorname{lcm}(\mu_1, \mu_2)/\mu_2$. Then $Z(s_1, s_2) = \nu_1 a_1 - \nu_2 a_2 = \nu_1(a_1 - \mu_1) - \nu_2(a_2 - \mu_2) \in \mathcal{I}(S)$ and hence $t^S\big(Z(s_1, s_2)\big) = 0$. Since every value of t^S is attained by some reduction, this has shown that the first claim implies the second, and it is trivial that the second implies the third.

It only remains to show that the third claim is in fact (a‴). To that end, let $(t_{v_1,s_1}, \mu, t_{v_2,s_2})$ be a V-critical ambiguity of $T_1(S)$. Let $(\mu_i, a_i) = s_i$ and $\nu_i = v_i(1)$ for $i = 1, 2$. Then $\nu_1\mu_1 = \mu = \nu_2\mu_2$ and hence $\operatorname{lcm}(\mu_1, \mu_2)$ divides μ. However if $\kappa := \mu/\operatorname{lcm}(\mu_1, \mu_2) \neq 1$ then $(t_{v_1,s_1}, \mu, t_{v_2,s_2})$ would be a proper V-shadow of $(t_{v_1/\kappa,s_1}, \mu/\kappa, t_{v_2/\kappa,s_2})$, which by criticality is not the case. Similarly $\gcd(\mu_1, \mu_2) \neq 1$ since one would otherwise have $\nu_1 = \mu_2$ and $\nu_2 = \mu_1$, in which case $(t_{v_1,s_1}, \mu, t_{v_2,s_2})$ would be a montage with composition map $w(b_1, b_2) = b_1 b_2$. Finally $Z(s_1, s_2) = \nu_1 a_1 - \nu_2 a_2 = t_{v_1,s_1}(\mu) - t_{v_2,s_2}(\mu)$. □

Another applied specialisation of Theorem 8.4 would be to take $\mathcal{Y}$ to be a monoid on the form $X_1^\bullet \times X_2^*$, where $X_1^\bullet$ denotes the free *abelian* monoid generated by X_1. This can be used to formally justify Gröbner basis calculations in $\mathcal{R}\langle X_2\rangle$ where the given relations contain some set X_1 of commutative coefficients for which one doesn't want to fix the values, by making the calculations in $\mathcal{R}[X_1]\langle X_2\rangle$ instead.

There is of course always the possibility to work in $\mathcal{R}\langle X_1 \cup X_2\rangle$ and add relations to make elements of X_1 commute with everything else, but that can get unintuitive and impractical (especially if X_1 is large compared to X_2). Another possibility would be to make a transcendental field extension of $\mathcal{R}$ with the variables in X_1, but that would then make it formally questionable to specialise to a case where the coefficients satisfy some algebraic relation.

Corollary 8.6. *Let $\mathcal{R}$ be an associative and commutative ring with unit, let X_1 and X_2 be disjoint sets, let $\mathcal{M} = \mathcal{R}[X_1]\langle X_2\rangle$, and let $\mathcal{Y}$ be the monoid in $\mathcal{M}$ which is generated by $X_1 \cup X_2$. Write $X_1^\bullet$ for the abelian submonoid of $\mathcal{Y}$ which is generated by X_1 alone. Let $V = \{b \mapsto \kappa\mu b\nu\}_{\kappa\in X_1^\bullet,\mu,\nu\in X_2^*}$ (a set of maps $\mathcal{M} \longrightarrow \mathcal{M}$). Let P be a well-founded semigroup partial order on $\mathcal{Y}$.*

Let $S \subseteq \mathcal{Y} \times \mathcal{M}$ be such that $a \in \mathrm{DSM}(\mu, P)$ for every $(\mu, a) \in S$ and let $T_1(S)$ be as in Construction 7.2. Then the three claims that $\mathrm{Red}(S) = \mathcal{M}$, $\mathcal{M} = \mathrm{Irr}(S) \oplus \mathcal{I}(S)$, and $\{\mu - a\}_{(\mu,a)\in S}$ is a P-monic V-Gröbner basis of $\mathcal{I}(S)$ are each equivalent to the conjunction of the following two conditions:

- *For every octuplet $\big((\mu_1, a_1), (\mu_2, a_2), r_1, r_2, r_3, \nu_1, \nu_2, \nu_3\big) \in S^2 \times (X_1^\bullet)^3 \times (X_2^*)^3$ such that $\mu_1 = r_1 r_2 \nu_1 \nu_2$, $\mu_2 = r_2 r_3 \nu_2 \nu_3$, $\nu_1, \nu_2, \nu_3 \neq 1$, and $\gcd(r_1, r_2) = \gcd(r_2, r_3) = \gcd(r_1, r_3) = 1$, there exists some $t \in T(S)$ such that $t(r_3 a_1 \nu_3 - r_1 \nu_1 a_2) = 0$.*
- *For every octuplet $\big((\mu_1, a_1), (\mu_2, a_2), r_1, r_2, r_3, \nu_1, \nu_2, \nu_3\big) \in S^2 \times (X_1^\bullet)^3 \times (X_2^*)^3$ such that $\mu_1 = r_1 r_2 \nu_1 \nu_2 \nu_3$, $\mu_2 = r_2 r_3 \nu_2$, $(\mu_1, a_1) \neq (\mu_2, a_2)$, and $\gcd(r_1, r_2) = \gcd(r_2, r_3) = \gcd(r_1, r_3) = 1$, there exists some $t \in T(S)$ such that $t(r_3 a_1 - r_1 \nu_1 a_2 \nu_3) = 0$.*

Proof sketch. Same overall structure as in the proof of Corollary 8.5, only the identification of V-critical ambiguities needs to be revised. This splits into a noncommutative part for X_2^* which is the same as in Example 6.10 and a commutative part for $X_1^\bullet$ which is the same as in Corollary 8.5. □

Corollary 8.7 (Gerritzen [6])**.** *Let $\mathcal{R}$ be a field, let X be a set, let $\mathcal{Y}$ be the free magma $\mathrm{Mag}(X)$ on X, and let $\mathcal{M}$ be the free $\mathcal{R}$-module with basis $\mathcal{Y}$. Extend the multiplication on $\mathcal{Y}$ to $\mathcal{M}$ by bilinearity, so that $\mathcal{M}$ is the (nonunital) free nonassociative $\mathcal{R}$-algebra $\mathcal{R}\{X\}$ on X. Let V_1 be the set of all maps $\mathcal{M} \longrightarrow \mathcal{M} : b \mapsto \nu b$ and $\mathcal{M} \longrightarrow \mathcal{M} : b \mapsto b\nu$ for $\nu \in \mathcal{Y}$. Let V be the monoid (with composition as operation) generated by V_1.*

Let P be a well-founded total order on $\mathcal{Y}$ such that

$$\lambda < \mu \text{ in } P \quad \Longrightarrow \quad \lambda\nu < \mu\nu \text{ in } P \text{ and } \nu\lambda < \nu\mu \text{ in } P \tag{8.2}$$

for all $\lambda, \mu, \nu \in \mathcal{Y}$. Let $S \subseteq \mathcal{Y} \times \mathcal{M}$ be such that $a \in \mathrm{DSM}(\mu, P)$ for every $(\mu, a) \in S$ and let $T_1(S)$ be as in Construction 7.2. Then the following are equivalent:

1. *$\{\mu - a\}_{(\mu,a)\in S}$ is a P-monic V-Gröbner basis of $\mathcal{I}(S)$.*
2. *$\mathcal{M} = \mathrm{Irr}(S) \oplus \mathcal{I}(S)$.*
3. *For all $(\mu_1, a_1), (\mu_2, a_2) \in S$ and $v \in V$ such that $\mu_1 = v(\mu_2)$ there exists some $t \in T(S)$ such that $t\big(a_1 - v(a_2)\big) = 0$.*

Proof. It follows from (8.2) that all elements of V are strictly monotone with respect to P. Hence the conditions in Theorem 8.4 are fulfilled and one only has to verify that the last condition is (a$'''$) by characterising the V-critical ambiguities.

An arbitrary ambiguity of $T_1(S)$ has the form $(t_{v_1,(\mu_1,a_1)}, \mu, t_{v_2,(\mu_2,a_2)})$ where $v_1(\mu_1) = \mu = v_2(\mu_2)$. The situation in the last condition is exactly this for $v_1 = \mathrm{id}$ or $v_2 = \mathrm{id}$, so it only remains to show that all other ambiguities are non-V-critical. Unique factorisation in $\mathcal{Y}$ gives rise to a unique factorisation in V (as compositions of elements of V_1), and thus there exist $v_1', v_2' \in V_1$ and $v_1'', v_2'' \in V$ such that $v_1 = v_1' \circ v_1''$ and $v_2 = v_2' \circ v_2''$.

If $v_1' = v_2'$ then $\big(t_{v_1'',(\mu_1,a_1)}, v_1''(\mu_1), t_{v_2'',(\mu_2,a_2)}\big)$ is another ambiguity, of which $(t_{v_1,(\mu_1,a_1)}, \mu, t_{v_2,(\mu_2,a_2)})$ is a proper V-shadow. Since there for every $\mu \in \mathcal{Y}$ is only finitely many $(v, \nu) \in V \times \mathcal{Y}$ such that $\mu = v(\nu)$, it follows that V-shadow-critical is the same as V-shadow-minimal, and hence none of the ambiguities with $v_1' = v_2'$ are V-critical.

If instead $v_1' \neq v_2'$ then one of these must multiply on the left and the other must multiply on the right; it can without loss of generality be assumed that $v_1'(b) = \nu_1 b$ and $v_2'(b) = b\nu_2$. This implies that $\mu = \nu_1\nu_2 = v_2''(\mu_2)v_1''(\mu_1)$ however, and thus $(t_{v_1,(\mu_1,a_1)}, \mu, t_{v_2,(\mu_2,a_2)})$ is a montage of $(\nu_2, t_{v_1'',(\mu_1,a_1)})$ and $(\nu_1, t_{v_2'',(\mu_2,a_2)})$ with composition map $w(b_1, b_2) = b_2 b_1$. Hence the ambiguities with $v_1' \neq v_2'$ aren't V-critical either. □

8.3 Path algebras

There is in the literature also a theorem by Farkas, Feustel, and Green [5] which similarly characterises (reduced) Gröbner bases in path algebras and certain semigroup algebras; the result is derived in an axiomatic setting generalising path algebras. Not surprisingly, it is in that setting equally possible to derive from the generic diamond lemma theory a result on more general (uniform monic) Gröbner bases in these algebras. The proof is essentially the same as for Theorem 8.4, but the result is not Yet Another Corollary due to some technicalities caused by allowing the product of two monomials to be zero.

In the present notation, one is given a field $\mathcal{R}$ and an associative $\mathcal{R}$-algebra $\mathcal{M}$ with basis $\mathcal{Y}$. This basis is assumed to be well-ordered, so let P

be that order. Another binary relation *divides*, or symbolically |, is defined on $\mathcal{Y}$ by $\mu \mid \lambda$ iff there exist $\nu_1, \nu_2 \in \mathcal{Y}$ such that $\lambda = \nu_1 \mu \nu_2$. These data are furthermore required to satisfy five axioms:

M1. $\mathcal{Y} \cup \{0\} \subset \mathcal{M}$ is a semigroup under multiplication.

M2. 'Divides' is reflexive.

M3. For each $\lambda \in \mathcal{Y}$, the set $\{\, \mu \in \mathcal{Y} \,|\, \mu \text{ divides } \lambda \,\}$ is finite.

M4. If $\mu, \nu, \lambda, \rho \in \mathcal{Y}$ are such that none of the products below are zero, then

$$\nu < \mu \text{ in } P \quad \Longrightarrow \quad \lambda\nu\rho < \lambda\mu\rho \text{ in } P. \tag{8.3}$$

M5. If $\mu \mid \lambda$ then $\mu \leqslant \lambda$ in P.

In the case that $\mathcal{M}$ is the path algebra $\mathcal{R}\langle\Gamma\rangle$ and $\mathcal{Y}$ is the set of all paths[2] in Γ (counting vertices as paths of length 0), axioms M1–M3 are trivial properties; in particular M1 is characteristic. M4 is a natural modification of the monoid partial order axiom (6.5) and M5 is another condition on P; the authors suggest that one meets it by using a length-lexicographic order, although a weighted-degree lexicographic order will work just as well. It should be observed that $\mathcal{Y} \cup \{0\}$ is typically not a monoid, since the unit in a path algebra is the sum of all length 0 paths rather than any particular path.

Simple reductions may be constructed as in Construction 7.2, with V being the set of maps $b \mapsto \lambda b \rho$ for $\lambda, \rho \in \mathcal{Y}$; this is exactly the same as in [5, p. 731]. Similarly the definition there of a (P-monic) 'Gröbner generating set' is exactly the same as 'P-monic V-Gröbner basis' here. Axiom M4 is exactly what is needed in Lemma 6.5 to establish that V correlates P to itself, and then the compatibility with P of $T(S)$ follows from Lemma 7.4 for any S constructed as in Lemma 8.3. It is however not quite as straightforward to apply Lemma 7.3 to prove that the elements of V are advanceable. Besides the trivial detail that V is not in general closed under composition —if $v_1(b) = \lambda_1 b \rho_1$ and $v_2(b) = \lambda_2 b \rho_2$ then $(v_1 \circ v_2)(b) = \lambda_1\lambda_2 b \rho_2\rho_1$ which is only an element of V if $\lambda_1\lambda_2 \neq 0$ and $\rho_2\rho_1 \neq 0$, although that can be worked around by considering $V \cup \{0\}$ instead— there is in this lemma also the more significant condition that every $(\mu, a) \in S$ and $v \in V$ must satisfy $v(a) = 0$ if $v(\mu) = 0$. This is why the result was above described as being about *uniform* monic Gröbner bases.

In [5, p. 733], two elements $\mu, \nu \in \mathcal{Y}$ are defined to be *uniform-equivalent* if

$$\lambda\mu\rho = 0 \quad \Longleftrightarrow \quad \lambda\nu\rho = 0 \qquad \text{for all } \lambda, \rho \in \mathcal{Y}. \tag{8.4}$$

[2] To be formally correct, one should really say *walk* rather than 'path', since a *path* (as all graph theorists know) may not have any repeated vertices, but speaking of 'walk algebras' here would probably cause more confusion than it avoids.

In a path algebra, this simply means that μ and ν have the same endpoints, but in principle the matter might be more complicated. Nonetheless, uniform-equivalence is an equivalence relation on $\mathcal{Y}$ and defines a partition of $\mathcal{Y}$ into equivalence classes. An element a of $\mathcal{M}$ is said to be *uniform* if all elements of $\mathrm{supp}(a)$ are uniform-equivalent, and consequently a pair $(\mu, a) \in \mathcal{Y} \times \mathcal{M}$ can be said to be uniform if every element of $\mathrm{supp}(a)$ is uniform-equivalent to μ. Considering only uniform Gröbner bases may seem like a severe restriction, but at least in the case of a path algebra it is actually rather trivial. The reason for this is that there is in a path algebra no way in which a path can be "uniform-superior" to another path; they're either equivalent or quite different. More concretely, if $\mu, \nu \in \mathcal{Y}$ are *not* uniform-equivalent then for each $v \in V$, at most one of $v(\mu)$ and $v(\nu)$ can be nonzero. This has the effect that only the uniform parts of rules get encoded into $T_1(S)$; for $t_{v,(\mu,a)}$ to even exist $v(\mu)$ must be nonzero and thus all $\nu \in \mathrm{supp}(a)$ which are not uniform-equivalent to μ will be killed by v.

In a path algebra, it is easy to see that any ideal is generated by a set of uniform elements; writing Γ_0 for the set of vertices in Γ, any $a \in \mathcal{R}\langle\Gamma\rangle$ can be expressed as the sum of uniform elements $\sum_{\kappa,\rho\in\Gamma_0} \kappa a \rho$, and these terms are elements of every ideal containing a. That the same should hold in general is not obvious, but any algebra satisfying M1–M5 must contain idempotent elements which fill the role of vertices in this argument; in particular axiom M2 is not as innocent as it may seem, since what it claims is really that there for every $\mu \in \mathcal{Y}$ exist $\kappa, \rho \in \mathcal{Y}$ such that $\kappa\mu\rho = \mu$. The structure of algebras satisfying M1–M5 is the subject of [5, Sec. 4], and the conclusion is roughly that any such algebra has to be a path algebra in which some paths have been identified.

Anyhow, with $\mathcal{M}$, $\mathcal{R}$, $\mathcal{Y}$, V, P, S, and $T_1(S)$ as above, it follows that (a), (a′), (a″), (a‴), (b), (c), and (d) of Theorem 8.4 are equivalent. (When employing Lemma 8.3 one must extend V with the identity map to make it a monoid, but since S is uniform that doesn't contribute any additional reductions.) The structure of V-critical ambiguities can be analysed as in Example 6.10; [5] gives the characterisation of overlaps between (μ_1, a_1) and (μ_2, a_2) as being determined by $\nu_1, \nu_2, \lambda \in \mathcal{Y}$ such that $\mu_1 = \nu_1\lambda$, $\mu_2 = \lambda\nu_2$, $\nu_2 \neq \mu_2$, and $\nu_2 \neq \mu_2$.

Acknowledgments

Part of the research reported herein was carried out in 2003–2004, when the author was a postdoc at the Mittag-Leffler institute, participating in the NOG Noncommutative Geometry programme.

References

[1] F. Baader and T. Nipkow: *Term rewriting and all that*, Cambridge University Press, 1998; ISBN 0-521-45520-0 and 0-521-77920-0.

[2] G. M. Bergman: *The Diamond Lemma for Ring Theory*, Adv. Math. **29** (1978), 178–218.

[3] L. A. Bokut′: *Embeddings into simple associative algebras* (Russian), Algebra i Logika **15**, no. 2 (1976), pp. 117–142 and 245. English translation in Algebra and Logic, pp. 73–90.

[4] B. Buchberger: *Ein Algorithmus zum Auffinden der Basiselemente der Restklassenringes nach einem nulldimensionalen Polynomideal* (German: An Algorithm for Finding a Basis for the Residue Class Ring of a Zero-Dimensional Polynomial Ideal), Doctoral Dissertation, University of Innsbruck, Institute for Mathematics, 1965.

[5] D. R. Farkas, C. D. Feustel, and E. L. Green: Synergy in the theories of Gröbner bases and path algebras, *Can. J. Math.* vol. **45** (4), 1993, 727–739.

[6] L. Gerritzen: Tree polynomials and non-associative Gröbner bases, *J. Symb. Comp.* **41** (2006), 297–316.

[7] L. Hellström: *The Diamond Lemma for Power Series Algebras* (doctorate thesis), 2002, xviii+228 pp.; ISBN 91-7305-327-9; HTTP://`abel.math.umu.se/~lars/diamond/thesis.pdf` or ditto `/thesis.ps.gz`.

[8] L. Hellström: *A Rewriting Approach to Graph Invariants*, ⟨AGMF2 proceedings⟩, 2006. Also at HTTP://`abel.math.umu.se/~lars/diamond/paper-gr.pdf`.

[9] D. E. Knuth and P. B. Bendix: *Simple word problems in universal algebras*, pp. 263–297 in: *Computational Problems in Abstract Algebra (Proc. Conf., Oxford, 1967)* (ed. by J. Leech), Pergamon, Oxford, 1970. Reprinted as pp. 342–376 in *Automation of Reasoning Vol. 2* (ed. by J. H. Siekmann and G. Wrightson), Springer, 1983; ISBN 3-540-12044-0.

[10] S. MacLane: *Categorical Algebra*, Bull. Amer. Math. Soc. **71** (1965), 40–106.

[11] T. Mora: *Seven variations on standard bases*, preprint **45** (1988), Dip. Mat. Genova, 81 pp. Available for download on prof. Mora's home page, at HTTP://`www.disi.unige.it/person/MoraF/publications.html`. Also item 1082 in the RICAM Gröbner Bases Bibliography.

[12] M. H. A. Newman: *On theories with a combinatorial definition of "equivalence"*, Ann. of Math. **43** (1942), 223–243.

[13] L. Robbiano: *On the theory of graded structures*, J. Symbolic Comput. **2** (1986), no. 2, 139–170.

[14] A. I. Shirshov: *Some algorithmic problems for Lie algebras* (Russian), Sibirsk. Mat. Zh. **2** (1962), 291–296.

[15] W. T. Trotter: *Combinatorics and partially ordered sets*, Johns Hopkins University Press, Baltimore, 1992; ISBN 0-8018-4425-8.

Index